STUDIES ON THE PERCEPTION OF ORGANIC CHEMICAL STRUCTURES

Natalie Claire Kellett

A thesis submitted in part fulfillment of the requirements for the degree of Doctor of Philosophy in the Faculty of Science, at the University of Glasgow.

May 1978

ACKNOWLEDGEMENTS

I am indebted to many people for their assistance in completing this thesis. In particular, I thank most sincerely:

Among my colleagues at the University of Glasgow,

. Dr A. Johnstone, whose help, advice and criticisms have been a constant source of inspiration - even 12,000 miles furth of Glasgow;

. Professor D. Sharp, for his helpful criticisms;

. Mrs K. McIntosh, (formerly Miss K. Urquhart) for her assistance with the interviews, and her many other valued contributions.

. Mr T. Howe, Mr I. Duncan and Dr F. Garforth, for permission to use their results.

The members of staff and pupils of the many schools that participated in this project, without whose interest and assistance the project would not have been possible. I am particularly grateful to the members of staff, who so willingly undertook the administration of the tests.

Among my colleagues at Monash University, Melbourne,

. Professor P. Fensham, for his friendly welcome and guidance;

. Dr L. West, for many interesting and helpful discussions.

I also acknowledge with gratitude financial assistance from the Scottish Education Department, and the Social Science Research Council.

Finally, I would like to thank my family - Dr B. Kellett, for his encouragement and his invaluable criticism, and John and Ruth Kellett, for their patience and helpfulness.

Natalie C. Kellett

Melbourne, 1978

CONTENTS

SUMMARY

As part of an evaluation of the Alternative Chemistry syllabus, introduced into Scottish Secondary Schools in 1962, an investigation is made of learning difficulties experienced in one area of Organic Chemistry, "Condensation, Hydrolysis and Esterification reactions".

An experiment performed in 1973 confirms earlier reports of poor pupil performance in tasks related to this topic.

Two hypotheses are proposed, on the basis of observation of classwork and discussion with teachers and pupils. These suggest that pupils' learning difficulties may (a) have arisen because the pattern characteristics of extended structural formulae cause perceptual confusion, or may (b) be conceptual in origin. In particular, it is suggested that pupils' conceptual understanding of functional groups is inadequate.

Two series of experiments, forming a critical test of the hypotheses, are described. The first involves the use of a series of immediate recall tasks. The pretesting of these materials and the validation of the associated scoring systems, carried out in 1973, is reported. The results obtained when these tests were administered to a representative sample of 1292 Scottish Secondary pupils are presented. 210 'H' Grade pupils participated in the second set of experiments. Measurements of pupils' ability to identify members of organic families are given. The results of a small scale interview are reported, and a comparison is made of pupils' performance on the identification tasks before and after using learning materials designed to lessen the effect of inadequate conceptual understanding.

The combined experimental results indicate that pupils are not confused by the pattern characteristics of formulae, but that less than 10% of pupils (even at Sixth Year Studies level) have achieved an adequate level of conceptual understanding of functional groups.

A third hypothesis is proposed, relating information content of tasks, their adjudged difficulty and pupils' conceptual understanding. Evidence consistent with this hypothesis is found in an analysis of the results of three independent studies.

Recommendations are made concerning the teaching of Organic Chemistry, and the further testing of this hypothesis.

INTRODUCTION

This thesis reports an investigation into the difficulties Scottish Secondary pupils have reported in learning certain sections of the Organic Chemistry courses specified in the new Alternative Chemistry syllabuses - in particular, the topic of "Condensation, Hydrolysis and Esterification Reactions".

The intention underlying the design of each phase of this investigation has been to build up a theoretical understanding of the cause(s) of these learning difficulties. This was not done by imposing any a priori theoretical framework upon the investigation. Rather, as the problem had arisen in the classroom situation, we looked to the classroom to provide a basis for the answer. Following direct observation of class work and discussions with teachers and pupils, certain causes of the observed learning difficulties were postulated. These hypotheses were then subjected to specific, critical testing. In following this approach we have endeavoured to meet the three criteria of relevance, adequate conceptualisation and appropriate methodology recommended by Tyler(1) in his analysis of research in Science Education (in which he considers the evaluation of new curricula in some detail).

We have sought a theoretical understanding of the causes of the learning difficulties, rather than, say, a delineation of the mistakes pupils make in this area, for two reasons. First, we believe that such an understanding will provide the best basis for determining the measures necessary to overcome these particular difficulties. Secondly, as Tyler(1) has said,

> "The object of research is generalisation; that is, the discovery of or the formulation of something which has wider applicability than a description of the particular case or cases which were the subjects of the study".

A problem area such as this one represents a failure of the normal learning process. A theoretical explanation of one such failure is at least potentially generalisable to other areas of learning difficulty, and may also offer the possibility of increasing our understanding of the processes that are involved in the successful learning of Chemistry. It is much less likely that a list of specific mistakes would allow such generalisations.

No single body of research can be seen as the precursor of this study. Rather, as the investigation proceeded, we found several separate areas of research relevant to our problem. As their relevance will be more evident in context, we do not attempt to discuss these topics here, but will introduce and review them at appropriate points in the text.

The immediate background to the present work is presented in Chapter 1, where we discuss Johnstone's original survey that initiated his ongoing evaluation of the Alternative Chemistry Syllabuses used in Scottish Secondary Schools. We also report a replication of his findings concerning pupils' performance in tasks related to Condensation, Hydrolysis and Esterification reactions.

The second Chapter contains a report of the observations and discussions that lead to the proposal of two hypotheses. The first - the "Visual Difficulties" hypothesis - proposed that pupils' difficulties arose because they were confused in some way by the pattern characteristics of the extended structural formulae used to represent organic compounds. The second hypothesis proposed that the difficulties were conceptual in origin, (and that, consequently, pupils are unable to extract or interpret the chemical information content of these formulae). In formulating a set of testable questions, we were led to consider the processes involved in the learning of science concepts. The review of the literature, reported in this Chapter, suggested that commonly used terms (such as "concept acquisition") were not altogether appropriate in this particular context. We have therefore proposed a new terminology, which we believe to be useful and appropriate in describing the learning of science concepts.

Operationally speaking, to test the Visual Difficulties and Conceptual Difficulties hypotheses, we had to determine how pupils perceive extended structural formulae. Because of its central importance, the study as a whole has taken its name from this procedure. The test materials used for this purpose consisted of a series of immediate recall tasks. The design of this test instrument, and the validation of the associated scoring systems, are reported in Chapter 3. At this time too, we first consider the relationship between Short Term Memory (or Working Memory) and pupils' learning difficulties; a relationship which is considered further, and in more detail, in Chapters 4 and 5.

Chapter 4 presents and discusses the results obtained when the tests were administered to a large sample of Scottish Secondary pupils, representative of the 'O' Grade, 'H' Grade and Sixth Year Studies populations.

A second series of experiments was designed to provide an additional test of the Conceptual Difficulties hypothesis. These are described in Chapter 5, along with the results obtained when they were administered to a representative sample of 'H' Grade pupils.

Taken together, the results of the two sets of experiments contra-indicated the Visual Difficulties hypothesis, and supported the Conceptual Difficulties hypothesis. This led to a very interesting situation. On the basis of the latter hypothesis, one would have expected widespread learning difficulties within Organic Chemistry, whereas pupils had reported difficulty in only certain Organic topics. Therefore, at the conclusion of Chapter 5 we propose a third hypothesis - the I.C.C.U.D. hypothesis - which could explain the observed selectivity. This hypothesis postulates a relationship between the information content of tasks characteristic of a topic, the apparent or rated difficulty of that topic, and pupils' levels of relevant conceptual understanding. An important aspect of this hypothesis is that the variables it considered, and the relationships it describes, are in no way specific to Condensation, Hydrolysis and Esterification Reactions.

While we have not attempted to validate this hypothesis in the present study, we have looked for evidence consistent with it in the results of three independent studies in different areas of Chemistry. This analysis is presented and discussed in Chapter 6.

The final Chapter reviews the study as a whole, and argues the case for a critical test of the I.C.C.U.D. hypothesis. On the basis of our theoretical findings, some specific recommendations are made for the teaching of Organic Chemistry.

CHAPTER 1

Condensation, Hydrolysis and Esterification Reactions - An Area of Difficulty

1.1 The First Report of Difficulty in Learning Organic Chemistry

In 1962 the new, 'Alternative' Chemistry syllabuses were introduced in Secondary Schools in Scotland.[(2)] Experienced chemistry teachers had been closely involved in the production of these syllabuses, and certainly they were enthusiastically received by teachers. However, a syllabus must ultimately be judged not on the basis of teachers' preference for it, but on the basis of whether or not it 'works' for learners. If it is to be successful for learners, the stated objectives for each level of the course (independent of their educational merit) must be attainable in the classroom situation by pupils of the corresponding age, and range in level of maturity and intellectual development. Furthermore, the syllabus content, and its ordering must enable learners to attain such objectives.

It is a very difficult matter to assess on purely theoretical grounds the success of a syllabus. For instance, even if we can correctly classify each topic in terms of the required Piaget level of development,[(3)] as Ingle and Shayer[(4)] have attempted to do for the Nuffield 'O' Level Chemistry course, there still remains an uncertainty in the average age of attainment of the Stage 3 formal operations so necessary to the mastery of many topics in Chemistry. Piaget and Inhelder have suggested that these skills are evident at 11-12 years, with full development at 14-15 years. Shayer[(5)] has suggested that the latter age might represent the beginning of Stage 3 development for the average British pupil. Lovell[(6)] reported a series of studies that suggested the onset of formal thinking was task-dependent, and Dale,[7] in a replication study of one of Piaget's chemistry experiments, suggested that there 'was no sharp transition from concrete to formal thinking at age 11-12 years. It appears that there is a gradual ... increase in ability to solve this problem.' He reported that the ability to solve the problem appeared at age 10, and was still increasing at age 15 years. These results suggest that there must be an inherent uncertainty in the best theoretical ordering, on psychological grounds, of topics within a chemistry syllabus. Again, it may be that the logical ordering of certain course material is very evident to a trained chemist, but this logic may not be at all apparent to someone actually learning the material; for him, a rather different order may represent the simplest and most effective route to mastery. Thus, however successful a syllabus may appear 'on paper' we must, for the present at least, rely heavily on empirical investigations to determine its effectiveness.

It was for this reason that A.H. Johnstone began, in 1969, an experimental evaluation of the working of the Alternative Chemistry syllabus in Scotland.[8] One aim of this investigation was the identification of any areas in which the syllabuses were not functioning adequately.

In an early phase of this investigation, Johnstone obtained, by questionnaire, pupils' assessments of the difficulty of topics in the 'Alternative' Chemistry course they had just completed. Pupils were asked to give one of four responses for each topic:

(a) 'easy to grasp' - i.e. understood with little effort when the topic was first taught,

(b) 'difficult to grasp' - i.e. required considerable effort to understand the topic,

(c) 'never grasped' - i.e. the topic was not understood and would have to be retaught.

(d) never taught.

The 'O' grade, 'H' grade and Sixth Year Studies syllabuses were assessed; for each, a large sample of pupils in two consecutive years was surveyed. In each case, the two sets of results showed a very high degree of coincidence in pupils' assessment of the difficulty of course topics. The results of the 'O' grade survey showed that certain groups of topics were judged to be very difficult by a substantial proportion of pupils. A group of topics, concerned with various condensation, hydrolysis and esterification reactions, was one of these 'areas of difficulty'. To illustrate the sort of responses obtained, the results relating to these organic topics obtained from one of the 'O' grade surveys(9) are reproduced in Table 1.1. The responses for topic G1 - an easy topic - are included for comparison.

TABLE 1.1

PUPILS' REACTIONS TO SOME 'O' GRADE ORGANIC TOPICS

RESPONSES ARE EXPRESSED AS PERCENTAGES

Topic		Easy to Grasp	Difficult to Grasp	Never Grasped
L3	The formation of Addition polymers e.g. perspex, polystyrene, pvc.	51	36	13
N1	The breaking down (hydrolysis) of carbohydrates using saliva or hydrochloric acid.	48	43	9
N2	The formation of Esters.	27	52	21
N3	The conversion of fats to soaps.	39	45	16
O1	The formation of condensation polymers e.g. nylon, phenol-formaldehyde.	32	47	21
G1	Atomic particles and their arrangement in the atom.	83	16	1

Organic topics - again, condensation and hydrolysis reactions, together with the descriptive chemistry of newly introduced families - were assessed as very difficult by pupils in the 'H' grade survey. The Sixth Year Studies pupils were asked to rate course topics and also certain concepts. Organic work generally was assessed as one of the more difficult areas of the course. The assessment of course concepts is shown in Table 1.2; the organic concepts are marked with an asterisk.

TABLE 1.2

REACTIONS TO SOME CONCEPTS IN THE SIXTH YEAR STUDIES COURSE

RESPONSES ARE EXPRESSED AS PERCENTAGES

Concept	Easy	Difficult	Never Grasped
Free Energy change (G)	61	35	4
Entropy change (S)	69	28	3
The mole	82	16	2
Absorption spectra	68	28	4
Orbitals	53	42	5
S_N1 and S_N2 reactions*	38	47	15
Grignard reactions*	42	44	14
pH and buffers	60	34	6
Origin of colour	59	32	9
Orbitals (degenerate and split)	37	48	15
Paramagnetism	60	31	9

These responses show that the two organic concepts listed were sources of real difficulty in the opinion of these students.

1.2 Delineation of the Research Area for this Project

The picture of learning in Organic Chemistry that emerged from Johnstone's survey was that, at each level, a substantial proportion of pupils (even the successful ones) were reporting difficulty with any new family or new type of reaction that was introduced. The fact that new organic work was rated as difficult even by the Sixth Year Studies pupils suggested that the problem might not be one of maturity. "Calculations involving the mole" was a topic reported as very difficult in both the 'O' grade and 'H' grade surveys; however, the figures shown for 'the mole' in Table 1.2 indicated that this was no longer a problem for Sixth Year Studies pupils. This is the trend one would expect to observe if an important concept or group of concepts is introduced to pupils at an age when they are not able to cope with the required degree of abstraction or complexity.

The results of the survey were consistent with the possibility that, for some reason, pupils were able to learn Organic Chemistry only inefficiently or ineffectively. For example, 'O' grade pupils who had found esterification a difficult topic, might feel their learning to be more successful when they met the topic again in their 'H' grade course. Another possibility was that difficulties experienced in learning organic topics at 'O' grade level might have carried over into subsequent years, either in terms of some underlying confusion detrimental to later learning, or in terms of a negative attitude to further organic work.

So that none of these possibilities should be excluded, it was decided that an investigation of the difficulties reported in the learning of Organic Chemistry should be concentrated, in the first instance, on those organic topics which were reported as difficult in the 'O' grade survey, and which were also studied in subsequent years.

Thus, the research project which is reported in the first part of this work, was defined as "An Investigation to Identify the Factors responsible for the difficulties experienced in the Learning of the Topics of Condensation, Hydrolysis, and Esterification Reactions". This investigation was to cover the learning of these topics in the 'O' grade, 'H' grade and Sixth Year Studies courses, and it was hoped that methods of overcoming these difficulties could be proposed.

1.3 Independent Verification of the Survey Results

Johnstone[9] had reported results obtained from six objective tests (pretests of items for subsequent '0' grade examinations) which had been administered independently of his survey, and which covered some of the topics that appeared in the '0' grade survey. The sample used for these tests was representative of the whole '0' grade population. The results of the tests that related to Organic topics are reproduced in Table 1.3, together with the difficulty rating assigned to each topic (in the survey). The results for topic G1 are again included for comparison.

TABLE 1.3

THE RESULTS OF '0' GRADE STUDENTS IN OBJECTIVE TESTS AND THE REPORTED DIFFICULTY RATING FOR CERTAIN TOPICS

Topic	Number of Items	Students' Reaction (Questionnaire)	% Giving Correct Response	
L3	4	Difficult	45,54,24,50	Av.43
N1	2	Inconclusive	51,53	Av.52
N2	3	Difficult	30,31,32	Av.31
G1	4	Easy	69,58,63,89	Av.77

These results suggested that the responses pupils had given in the '0' grade survey questionnaire were accurately reflecting a genuine difficulty in learning.

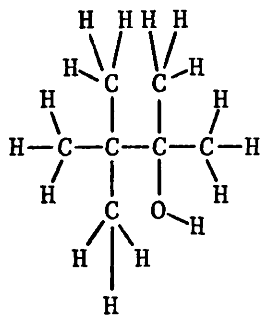
When the present investigation was commenced early in 1973, these '0' grade results provided the only quantitative published measure of pupils'performance in Organic topics.

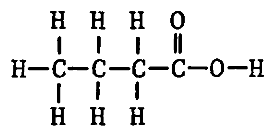
In order to make a comparison between task performance and difficulty rating at both 'O' grade and 'H' grade levels, a short experiment was conducted during the Open Day at Glasgow University in October 1973. Groups of pupils who had just commenced fifth or sixth year chemistry courses (and who had therefore successfully completed 'O' grade courses and 'H' grade courses, respectively, in June) were asked to participate in the experiment during their visit to the Chemical Education exhibit. Conducting the experiment under these conditions imposed certain restrictions on the test design, and on the type of information that could be obtained. First, it was not possible to arrange that a representative sample of the 'O' grade and 'H' grade populations should take part in the experiment. However, study of the lists of schools that had visited the Open Day in previous years showed that a wide variety of schools could be expected to attend. Therefore, it seemed reasonable to describe the sample of pupils who participated in the experiment as varied, if not necessarily representative. Secondly, the test had to be administered in 10-15 minutes. This limited the tasks to be used to those that could be completed quickly. Thirdly, as pupils would not have studied any organic Chemistry since the previous June, the test performance had to depend minimally on straight recall.

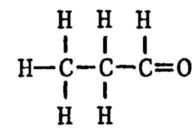
One learning objective of both the 'O' grade and 'H' grade syllabuses was the ability to identify correctly the family to which a given compound belonged. This skill incorporates the subsidiary skill of being able to identify a functional group correctly. The test used examined the two skills independently. The tasks chosen could be performed quickly by pupils. Associating the appropriate family name with a family member may or may not involve pure recall; however, remembering the family names themselves certainly involves only recall.

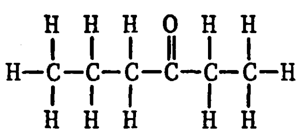
CHEMICAL EDUCATION: OPEN DAY EXPERIMENT

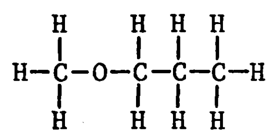
1. Put a ring around the part(s) of each molecule that you think most important in determining the behaviour of that compound.

(i) 

(ii) 

(iii) 

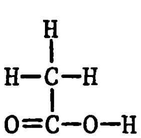
(iv) 

(v) 

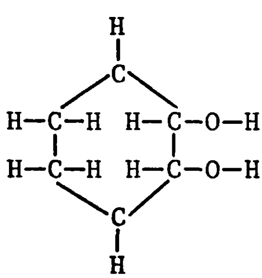
2. What kind of compound is each of the following? Put the appropriate letter below each formula.

CODE

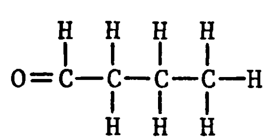
A = Aldehyde
B = Carboxylic Acid
C = Ether
D = Ketone
E = Ester
F = Alcohol
G = Carbohydrate
H = Aromatic



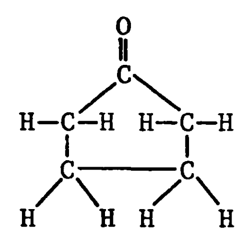
1. ______



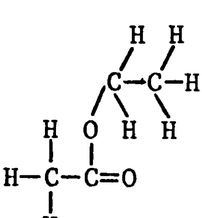
2. ______



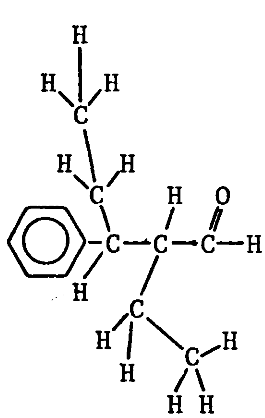
3. ______



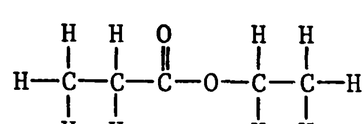
4. ______



5. ______



6. ______



7. ______

Fig 1.1 Question Sheet for Open Day Experiment.

To fulfil the requirement that task performance should depend minimally on recall, pupils were given a list of family names, and had only to identify the correct name from this list.

1.31 <u>Test Procedure</u>

Each group of fifth or sixth year pupils that volunteered to participate in the experiment was divided into two. Approximately two thirds of the pupils were asked to complete the question sheet which is reproduced in Fig. 1.1. The remaining pupils were shown five models of organic molecules (two carboxylic acids, two ketones, an ester and an alcohol). Their response sheets contained the list of family names shown in Question 2, Fig. 1.1, and a listing of the colour of the sphere that represented each element used in the models. These pupils were asked to indicate what type of compound was represented by each model.

All pupils were asked to indicate which year of chemistry they were studying. Unfortunately, half the sample failed to give this information, so it was not possible to analyse the fifth and sixth year responses separately. Pupils who participated in the molecular model question were asked to indicate whether they used such models occasionally/regularly/never in school work. As almost all pupils replied 'occasionally' it was not possible to make the intended comparison of task performance with level of use of models during learning.

1.32 <u>Test Results</u>

1. Question 1, Fig.1.1

This question tested (indirectly) pupils' ability to identify functional groups correctly. 100 pupils answered this question sheet. The correct responses (as percentages) for items (i)-(iv) are shown in Table 1.4. Item (v), an ether, was included as an aid to identifying those pupils who simply ringed the end of a compound. The figures shown for compounds (ii) and (iii) do not include the responses of 12 pupils who had ringed the functional groups correctly, but whose responses overall consisted of rings around the end of each molecule. The responses of a further 15 students suggested the same tendency, but not sufficiently strongly to warrant marking them incorrect.

TABLE 1.4

CORRECT RESPONSES FOR QUESTION 1 (N = 100)

	Alcohol	Acid	Aldehyde	Ketone
Correct Responses	55	51	53	77

2. Identification of Family, given Structural Formulae. Question 2, Fig. 1.1

Twelve pupils who completed Question 1 gave no responses at all to this question. It is unlikely that they had insufficient time to complete this question, but nevertheless they were excluded from the analysis of the responses for Question 2. The responses of the remaining 88 pupils are tabulated, as percentages, in Table 1.5.

3. Identification of Family, given Molecular Models

61 pupils attempted this question. The responses they gave are tabulated as percentages in Table 1.6.

1.33 Discussion of Results

The figures in Table 1.4 indicated a poor level of performance in items (i)-(iii). The most frequent mistake made was the choice of only part of a functional group, apart from the rather indiscriminate choice

TABLE 1.5

RESPONSES ARE GIVEN AS PERCENTAGES (N = 88)

STUDENT RESPONSES

Identification of Family Members

Formulae	Labelled Only Positive Instance(s)	Also Labelled Negative Instance(s)	Labelled Compound(s) Incorrectly	No Label Given	% Students Not Showing Success
Acid (Item 1)	46.6	5.7	29.5	18.2	53.4
Aldehyde					
(3) only	22.7	2.3	19.3	13.6	
(6) only	3.4	1.1	34.1	19.3	
both	1.1	0	25.0	15.9	99.0
Alcohol (2)	29.5	1.1	39.8	29.5	70.5
Ester					
(5) only	18.2	2.3	12.5	12.5	
(7) only	8.0	0	21.6	15.9	
both	2.3	0	17.0	35.2	97.7
Ketone (4)	8.0	4.5	31.8	55.7	92.0

TABLE 1.6

RESPONSES ARE GIVEN AS PERCENTAGES (N = 61)

Models	Labelled Only Positive Instance(s)	Also Labelled Negative Instance(s)	Labelled Compound(s) Incorrectly	No Label Given	% Students Not Showing Success
Acid					
(1) only	19.7	0	26.2	11.5	
(3) only	23.0	1.6	26.2	8.2	
both	4.9	1.6	32.8	1.6	95.1
Ester (2)	18.0	0	62.3	19.7	82.0
Ketone					
(4) only	8.2	1.6	24.6	4.9	
(6) only	16.4	3.3	9.8	9.8	
both	1.6	0	36.1	23.0	98.4
Alcohol (5)	44.3	4.9	34.4	16.4	55.7

of the end of a molecule, as mentioned above.

In interpreting the figures in Tables 1.5 and 1.6, allowance must be made for chance guessing of positive instances. As eight family names were given, any figure in Column 1 that is not significantly different from 12.5% may indicate chance guessing. For the sample sizes involved, the "guessing interval" (the interval of percentages not significantly different from 12.5%)[10] ranges from 0% to about 28%. Therefore, where a figure in Column 1 is less than 28%, we cannot reject the possibility that "identification" is due to guessing. On this basis, it would seem that only families for which non-chance response was a real possibility were the alcohol, and possibly acid families. A more detailed analysis of the results supports the view that there was a lack of consistency, or certainty, in students' responses.

A pupil was judged to have demonstrated the ability to identify a family member correctly if he labelled all positive instances of that family correctly, and did not give the family name to any negative instances. The figures in Column 5 of Tables 1.5 and 1.6 indicated that the performance in identifying members of individual families was very poor. A measure of the overall ability to identify members of families was obtained by calculating the mean number of 'families' correctly identified by each pupil. The mean for pupils who were given formulae was 0.88, and for those given models 0.70, (out of a possible 5).

The figure of 99% in Column 5 of Table 1.5 may be an overestimate of those that could not identify an aldehyde. Pupils' responses suggested that the inclusion of 'Aromatic' in the list of family names had added an unintended source of confusion to the experiment. 9% of the sample labelled item (3) correctly, and labelled item (6) an aromatic compound. However, even if this figure is subtracted from the

99% reported, we are still left with a figure indicative of a very poor level of performance. Altogether, 25% of the sample labelled item (6) 'aromatic'.

There were only two other instances in which a significant number of pupils chose the same incorrect response - and both occurred for item (3) of Question 2. 18% of the sample labelled this an alcohol, and 13% a ketone. Given the timing of this experiment, confusion between an aldehyde and a ketone would not have represented a serious error; however this was the only instance in which this confusion was evident in either version of the compound identification task.

Overall, then, the pupils performed very poorly on these experimental tasks. Of course, it could be argued that because of the time at which the experiment was conducted this performance was not a reliable indicator of pupils' true ability. There was, however, one feature of the pupils' responses which suggested strongly that the mistakes they made arose from fundamental misconceptions and were not simply attributable to the time of the year. This feature was the number of instances in which pupils gave two different responses for two examples of the same family (e.g. the two esters in Question 2), and in which the same family name was given to examples of different families (e.g. an ester and an acid both labelled ketone). The percentages of pupils giving at least one such misrepresentation are shown in Table 1.7.

TABLE 1.7

RESPONSES ARE GIVEN AS PERCENTAGES

	Same Family Name Given to Examples of Different Families	Two Examples of the Same Family Assigned to Different Families
Pupils given formulae (N = 88)	22	83
Pupils given models (N = 61)	15	98

The pairs of different responses for given examples of the same family included several combinations - one instance correctly identified and one incorrectly identified (the entries in Column 2 of Tables 1.5 and 1.6); one example labelled and the other not; and also a number of cases in which two different incorrect labels were given. With two exceptions, the pupils who are listed in Column 4 of Tables 1.5 and 1.6 as having given incorrect responses for both instances of a family came into this category. Many pupils gave pairs of different responses for both of the families represented by two positive instances; the overall performance in this respect is best described by noting that 74% of the total 'family pairs' were given different responses. These results suggest strongly that pupils were performing poorly not because they were unable to associate the appropriate name with a family member, but rather because they were not able to apply an appropriate criterion to determine family membership.

The results of this survey, then, provided additional evidence of a correspondence between pupils' subjective assessment of topic difficulty and their performance in a task of basic importance in that topic.

CHAPTER 2

Possible Origins of the Learning Difficulties

2.1 Introduction

The poor performance of pupils in Organic items, described in Chapter 1, could scarcely have come as a surprise to many Chemistry teachers in Scotland. In past years the Principal Examiner's reports on performance in 'O' and 'H' grade examinations have listed Organic work, particularly the topics of Condensation, Hydrolysis and Esterification, as an area of weakness with depressing regularity. The reports have consistently drawn attention to specific problems - for instance, the inability to give the correct structural formula of a named compound, and the inverse operation of naming a compound given its structural formula; an inability to name or identify the reactants required to give a particular ester has been noted, as has the difficulty experienced in identifying the monomers from which a given polymer has been obtained.

Rather than treating each such weakness as an independent problem we began this investigation by attempting to identify any general, underlying factors which could give rise to the type of mistakes that had been noted. First, an examination was made of the course content. Two possible sources of confusion were noted, and these are described in Section 2.2. Secondly, a series of visits were made to a variety of schools. The teaching - and learning - of Organic topics was observed in both 'O' and 'H' grade classes. This also provided an opportunity for informal discussions of Organic Chemistry with pupils, and with members of staff. As a result of these observations and discussions, two hypotheses relating to the origin of the difficulties in learning were formulated. These are stated and discussed in Section 2.3.

Section 2.4 contains a discussion of the learning or acquisition of science concepts. This establishes a framework within which the learning of specific chemical concepts will be considered in this work. When the type of learning traditionally studied in a concept learning investigation is compared with the learning processes required in science, little overlap seems apparent. Because of this, it was felt that the use of technical terms, such as 'concept acquisition', commonly used in such investigations, could not be applied unambiguously to the learning of science concepts. Therefore, Section 2.4 also contains a definition of a term which will be used extensively in this work in considering the learning of science concepts.

The final section in this chapter considers a set of specific, testable questions which formed the basis for an experimental test of the hypotheses described in Section 2.3. The experimental design for the initial phase of this investigation is outlined.

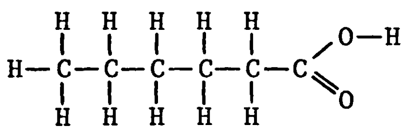
2.2 Course Content

When the results of Johnstone's survey were first published, many teachers were somewhat surprised at the very definite assessment of 'difficult to learn' that pupils recorded for C.H.E. topics. (i.e., Condensation, Hydrolysis and Esterification topics). The Organic topics happen to be listed at the end of the printed syllabuses for 'O' and 'H' grade chemistry, and, as the majority of teachers follow the syllabus order in course planning, Organic topics are generally the last to be taught. For this reason, it had been customary to see pupils' poor examination performance as a result of over hasty teaching or revision - or as a lack of extra revision by students who felt over confident of their ability to recall their most recent class work.

An examination of the Organic course content revealed a probable reason for teachers' surprise - from a chemist's point of view, the 'O' and 'H' grade syllabuses require pupils to learn very little. Generally speaking, the courses involve a simple descriptive study of the physical and chemical properties of the specified families, and a simple non-mechanistic description of the processes of condensation, hydrolysis and esterification. Wherever possible, reference is made to commonly occurring natural processes, (many of which would be studied by pupils in the biological sciences), and to important and widely used substances such as nylon. A great deal of practical work is also suggested for this section of the course. The various topics within the descriptive and C.H.E. sections form a chain of study rather than a pyramid of study. That is to say, there are no topics, which if not mastered initially, could give rise to the difficulties in learning of subsequent topics for which they form the foundations.

Only two sources of confusion became apparent in examining the course content. The syllabus suggests that pupils may be told that when a carboxylic acid and an alcohol react to form an ester and water, the acid provides the hydroxide units, and the alcohol the hydrogen, that ultimately form water molecules. Teachers are recommended not to discuss reaction pathways, but to treat this information as an 'interesting fact', and an example of the way in which Mass Spectroscopy may be used to elucidate a reaction process. Now, if a pupil had been taught that an acid was a substance that provided hydrogen (ions), and a base a substance that provided hydroxide ions, he might find the above description of esterification contradictory. It should be noted, however, that any confusion arising from this apparent contradiction could be expected to be related very specifically to esterification. The other source of confusion seemed to have more general implications. Memorandum paper No.3,[11] which gave detailed interpretations of the syllabus contents, recommended that five main types of organic formulae be used. The illustrative examples (all of hexanoic acid) given were:

C_3H_6O	$C_6H_{12}O_2$	$C_5H_{11}COOH$	$CH_3.CH_2.CH_2.CH_2.CH_2.C\begin{smallmatrix}OH\\ \nwarrow\\ O\end{smallmatrix}$
empirical formula	molecular formula	extended molecular formula	structural formula



extended structural formula

The first type of formula would normally be used in analytic work; junior pupils would use the second type in exercises related to isomeric compounds. The last three formula types, which give increasingly detailed information about the structure of a molecule, are all commonly used in descriptive Organic work. We felt that pupils could be confused (and that this confusion could lead to learning difficulties) if these three types of formula were used interchangeably, and particularly without explanation, during early phases of learning. A pupil might not be able to deduce for himself, that, for example, -COOH and $-C\begin{smallmatrix}\nearrow O\\ \searrow OH\end{smallmatrix}$ were equivalent representations of the same chemical entity. Particularly if a pupil relied on rote learning, he might recognize one of these representations as a salient cue for family identification, but not the other. If the arbitrary use of the three formula types had caused confusion, the result could be difficulty in learning in any or all of the descriptive Organic Chemistry topics, including C.H.E. reactions.

The examination of course content suggested that, during the visits to schools, three points that should be explicitly discussed with teachers were:

(i) Did they find the teaching of Organic Chemistry at 'O' and 'H' grades had to be hurried?

(ii) What definition of 'acid' would pupils have had prior to their 'O' grade Organic work?

(iii) What type(s) of formula were used in Organic work?

2.3 Teachers' and Learners' Points of View

Discussions with teachers indicated fairly clearly that the possible sources of confusion discussed in the previous section were not contributing in any major way to the learning difficulties. Although the vulnerable position of Organic work was acknowledged, all teachers felt that they had ample time to complete the syllabus. Some independent confirmation of this was obtained in subsequent years, when a representative sample of schools, that conducted experiments after the Organic topics had been taught, were able to return the results well before the end of the second academic term.

Because some of the schools visited were testing a revised order of teaching, different classes were using very different definitions of 'acid' at the time they commenced Organic Chemistry - e.g. some classes had been told only that an acid was a substance that turned blue litmus red. The way in which esterification was taught also varied from school to school. In some cases, the origin of the H and OH units was discussed in the fourth year, while in other schools this was not mentioned until the fifth year. However, in all these schools, similar difficulties in esterification exercises were observed and reported; this suggested that a link between 'acid' definition and learning difficulties associated with esterification was unlikely. Finally, teachers reported that they used extended structural formulae for junior descriptive Organic work, reverting to a more contracted form only in the fifth, or more frequently sixth year, work. Of course, it does not follow that all teachers, therefore, used such formulae. However, as learning difficulties had been apparent in these schools, where there was not the possibility of confusion due to the use of different formula types, it seemed unlikely that such confusion could be the major factor in causing learning difficulties.

All teachers with whom C.H.E. topics were discussed, agreed that pupils experienced marked learning difficulties in this area. At the same time, the consensus of opinion was that the topics should have been well within the competence of even fourth year pupils; it was felt that the material selected for study was appropriate in terms of cognitive content and general interest. Again, the practical work suggested was felt to be very suitable, and teachers made extensive use of experimental work (a great deal of which was carried out by pupils).

Only one suggestion was advanced regarding the source of the learning difficulties. Although this suggestion was made by many teachers, it generally seemed to be proposed as the only explanation that could be thought of, rather than as something that was manifestly true. The equations that are written to describe C.H.E. reactions have one special common characteristic; they all contain at least one formula written 'back-to-front'. The suggestion that teachers made was that pupils might be confused by seeing formulae written sometimes one way and sometimes the other. It was felt that this confusion could lead to learning difficulties, perhaps because pupils were unable to form a stable, accurate, internal representation of a functional group which they saw regularly in different orientations. (It was interesting that this suggestion in some ways paralleled the proposal that the use of different formula types could lead to confusion).

This suggestion merited critical testing both because of the weight of teaching experience supporting it and because of its prima facie plausibility. It was decided that the test should not be restricted to the one source of confusion mentioned by teachers. An extended structural formula purveys chemical information (to the initiated at least) but it is also a particular type of visual pattern, of which 'back-to-frontness' is only one characteristic. It seemed essential to allow for the fact that pupils could be hindered in their learning of Chemistry if they were confused by any characteristic of the visual formula-pattern. We therefore proposed that a critical test should be made of the hypothesis, 'That the learning difficulties are visual in origin.'

Discussion with pupils and observation of their work in class provided a rather different view of the problem. First, when back-to-front formulae were discussed, pupils said quite definitely and confidently that they were not difficult to use or to write. One group of pupils even demonstrated, very enthusiastically, a technique they had developed for themselves for writing formulae backwards. Now, it is quite possible for pupils to feel confident of their ability to perform a given task successfully, when, in fact, they perform it incorrectly. However, it is much less likely that a pupil could feel confident of his ability to handle a back-to-front formula and simultaneously be confused by it. To this extent, pupils' reactions argued against the suggested relationship between learning difficulties and that particular pattern aspect of a formula.

Branching or bending are also common pattern characteristics of formulae. In order to investigate pupils' reactions to these aspects of formula patterns, the identification of isomers was discussed. (This task would often involve the use of branched or bent formulae; and, as teachers had reported that pupils often performed poorly in this task, it seemed possible that a relation between learning difficulties and confusion due to the pattern aspects of formulae might be observed in this area). Pupils were shown two formulae, one of which was bent or branched, and were asked if they found it difficult to decide whether such formulae represented isomers. Most pupils felt that branching or bending did not present any problem, although different reasons were given for this attitude. The majority pointed out that it was often possible to 'straighten out' a formula, and then proceed to make the necessary comparison. However, some pupils stated that the bending or branching itself indicated that the substances (to quote) 'probably weren't isomers'. Further discussion revealed that a great many pupils believed, quite mistakenly, that two substances were isomers if their extended structural formulae were equivalent or, as some suggested, identical. (It is worth noting that a range of opinion existed as to what constituted 'equivalence' or 'identity'.) It seemed that these misconceptions alone could readily account for pupils' poor performance in isomer tasks, and certainly in this case, as in the case of back to front structures, pupils did not seem to be conscious of any confusion arising from the visual pattern characteristics of formulae per se.

Pupils did comment on one aspect of formulae, and that was their size. Again and again the complaint was made that big formulae were hard, and big equations were hard. Such statements were not amplified; sometimes they were made as general comments, while at others they were added to comments about back-to-front or bent structures as "that's all right, but they're hard if they're big." This seemed a most interesting point; it led to a proposal of a relationship between formula size and learning difficulties which will be discussed in a later chapter.

Although pupils did not themselves propose any causes of their learning difficulties, one possibility was suggested by many of their comments and their class work. These gave the definite impression that few pupils (from fourth year to Sixth Year Studies level) had gained, or were in the process of acquiring, the concept of a functional group. For instance, when pupils were introduced to esterification and were asked to suggest how an acid and alcohol might combine (after they had studied carboxylic acids and alcohols) they showed no expectation that the reaction would involve either functional group. Again, when pupils were asked why they thought, say, ethanoic acid and butanoic acid reacted in the same way, the common reply was that they belonged to the same family. No-one suggested that family membership and characteristic reactions were both determined by the functional group. Even Sixth Year Studies pupils were completely unable to offer any positive comments when asked whether they thought that carbonyl compounds (containing the $-\overset{O}{\overset{/\!/}{C}}-$ group) and carboxylic acids might be expected to show some similarity in behaviour, given that both contained a $-\overset{O}{\overset{/\!/}{C}}-$ unit. The only thought expressed here was that the compounds would be different because they belonged to different families.

It seemed reasonable to suppose that the lack of this concept could be related to learning difficulties in specific tasks. It also seemed that it could give rise to learning difficulties in a more general way. The concept of a functional group is a common thread linking the study of individual families - or specific functional groups - and also C.H.E. reactions. Without this common thread, the pupil would be left with a series of unrelated topics; each would have to be learnt individually and learning in one topic would not reinforce learning in other topics.

Also, if Organic Chemistry were not seen as the study of functional groups, the choice of particular families for study might seem arbitrary. Pupils, particularly in the fourth and fifth years, often asked, for instance, why were esters so important that they had to study them. Questions of this type seemed to imply that pupils felt each family must have been chosen because of its intrinsic practical (or commercial) importance - which was not obvious to them. Such an attitude would further exacerbate the fragmentation of learning, and could very possibly adversely affect motivation for learning.

So, as the teachers' point of view lead to the proposal of the hypothesis "That the learning difficulties were visual in origin", considering Organic Chemistry from the learners' point of view, we were led to propose a second hypothesis, "That the learning difficulties were conceptual in origin."

Before reformulating these two hypotheses in terms of operationally defined, testable questions, it is necessary to discuss what is meant by 'lacking' or 'having' a science concept.

2.4 The Acquisition of Concepts

The learning or acquiring of concepts is of major importance in Science Education. This is reflected on the one hand by the increasing number of 'modern' syllabuses and courses that have moved away from a 'traditional' rote-learning approach, toward an approach that emphasises the progressive understanding and application of principles and concepts.[12] Within the area of research, increasing interest has been shown in the application of Ausubel's theory of Learning to Science Education. Novak, Ring and Tamir[13] have reviewed this theory and its implications for Science Education research. An important element in Ausubel's theory is his distinction between "rote learning" and "meaningful learning."[14] In meaningful learning, new learning is incorporated in a non-arbitrary and substantive way into the learner's existing cognitive structure, while rote learning involves a purely arbitrary or piece-meal memorization of the new material. These two types of learning are not seen as mutually exclusive, but rather as the extremes of a learning continuum. In Ausubel's terms, meaningful learning can occur if the learner possess subsumers - that is, relevant, generalized ideas (or elements of cognitive structure) that allow new learning to be readily associated with and absorbed into existing structure. The realization that many science concepts and principles could play this subsuming role is one reason why Ausubel's theory appears so relevant for Science Education. West and Fensham,[15] in reviewing a number of early investigations of the way in which prior knowledge affected new learning in science, concluded that the existing evidence for the subsumption theory was indirect - they described it as 'evidence for the committed Ausubelean, but open to alternative explanations.' However, they also pointed out that the research evidence for Ausubel's theory 'was already stronger than that for some other processes (e.g. discovery learning) that have received acceptance in classroom practice.' In a recent article, West[16] has reported a study whose results strongly support the subsumption theory.

Thus, we have a situation in which the learning of particular science concepts is not only a desirable end in itself, but is also of potential use in the learning of new material.

Interest in concept formation has not, of course, been confined to Science Education. In considering the broad field of Concept Formation, we will make use of a categorization suggested by Vinacke[17] in his review of the subject. He suggested that, for experimental purposes, the problem of concept formation could be divided into three distinct areas, namely:

Area 1: The ability to conceptualise, which he amplified in the question "How can one explain and describe the development in the child of the ability to form and use a concept?"

Area 2: The acquisition of concepts, or repertory, that is "What concepts, or patterns of concepts, characterize various stages in the development of the child's thinking and acting?"

Area 3: Achieving a specific concept, that is, "What behaviour is manifest in attaining a particular concept, and what conditions influence that behaviour?"

Vinacke related the first two areas largely to the early stages of child development, and this age group is still heavily represented in research within these areas. A number of important investigations which fall within the description of Vinacke's Area 2 have been Science oriented; they have considered the order or complexity of concepts and the level of abstraction possible at different developmental stages, and many of these investigations have involved pupils of secondary school level. Lovell[6] has discussed recent work in this area, within the framework of Piaget's developmental theory of intellectual growth.

However, it is Vinacke's third area that is most pertinent to the present investigation. It is his amplifying question, "What behaviour is manifest in attaining a particular concept" which must be answered in order to operationally define the having or lacking of a particular science concept. Work within Area 3 has produced a large body of research findings, but there seems to have been little investigation specifically related to science concepts. The general point of view underlying investigations of concept achievement has produced a set of results and models that do not seem to provide directly an appropriate framework within which a formal description of the acquisition of a specific concept may be attempted.

At the time of Vinacke's review, most investigations within Area 3 had been conducted at the adult level. In fact, Vinacke specifically stated that, where adults were concerned, the third area was the relevant one because "the adult has already developed an ability to conceptualize, and has already acquired an enormous repertory of concepts." He added "It is probable, therefore, that the adult does not typically acquire new concepts, so much as he applies concepts which he already possesses, or learns new variations, hierarchies, etc., of these concepts."

More recent investigations have also generally involved adult subjects, and have been based on the same view of concept achievement. Typically, subjects are required to learn a class concept - that is, a concept which divides a series of stimulus patterns into a set of positive and a set of negative instances. During the course of an experiment, subjects are told whether each pattern is a positive or negative instance of the concept, and they must use this information to determine the concept chosen by the experimenter. The characteristics of the patterns such as size, colour, shape etc., are called 'dimensions', and in a particular experiment each dimension will take on a number of values (e.g. 'colour' could take on the values red, green and yellow). Each of these values is termed an 'attribute'. The concept chosen for an experiment consists of two or more attributes (the 'relevant' attributes), together with a combinatorial rule; for example, 'red and large' and 'red and/or large' are two concepts having the same relevant attributes, but differing combinatorial rules. In the first case, any large red pattern (irrespective of shape, which would be called an irrelevant attribute) would be a positive instance; in the latter case, the set of positive instances would contain all red patterns, all large patterns, and, of course, all large red patterns. It is worth noting here that the attributes selected - such as redness, squareness etc. - involve concepts that are certainly within the subjects' repertories; the experimental paradigm requires subjects to learn a new variation or grouping of such concepts.

The problem of learning or acquiring a concept, as defined by this sort of experiment, involves two tasks, namely, identifying the relevant attributes, and identifying the rule; these tasks are often studied independently. In attribute identification experiments, subjects will be informed of the rule that has been chosen, and conversely, in rule identification experiments, the relevant attributes will be specified.

Attribute learning investigations have considered such factors as the attention value of different kinds of cues or attributes,[18,19] and the way in which emphasis of a relevant cue affects its attention value;[20,21] the effect on learning of the numbers of relevant, irrelevant and redundant dimensions,[22,23,24,25] and of the number of values per dimension.[26]

In rule identification experiments, one finds investigations of the effect on learning of differing levels of rule complexity.[27,28,29] There have also been investigations of the distribution of learning between attribute identification and rule identification where these tasks are presented as simultaneous unknowns.[28,30]

During attribute or rule learning tasks, subjects must formulate and test hypotheses (e.g. 'red' is a relevant attribute). The sort of strategies used by subjects (that is, the number of hypotheses formulated, and the way in which they are tested) have been discussed by Bruner et al.[27] Various models of strategy selection have been proposed; for example, the role of memory in strategies has been considered,[25,31,32,33,34] and more recently attention has been given to strategies based upon truth table classifications.[35,36,37]

The processes involved in the type of concept acquisition considered above have been formally defined by Bruner et al.[38] as "concept formation" and "concept attainment." They have described "concept formation" as the formulation of an hypothesis (regarding the attributes) and "concept attainment" as the "process of finding predictive defining attributes that distinguish exemplars from non-exemplars of the class one seeks to discriminate." Concept formation is seen as a necessary first step for concept attainment.

Ausubel has also considered the acquisition of specific concepts, although from a point of view more broadly based than Bruner's. Ausubel, too, uses the term "concept formation"; his definitions of this term include the processes Bruner described separately as the sequence of concept formation and concept attainment. Thus, there is a great deal of similarity between Ausubel's "concept formation" and Bruner's "concept attainment." Ausubel[39] considers that concept formation is "characteristic of the pre-school child's inductive and spontaneous acquisition of generic ideas," but that it is also exhibited, at a more sophisticated level, by adults. In his detailed description of the processes involved in concept formation,[40] he includes the formulation and testing of hypotheses regarding the attributes, and the selection of a set of predictive defining attributes. As Ausubel considers the acquisition of concepts within the context of meaningful learning, he also proposes that the relation of the defining attributes to relevant anchoring ideas in cognitive structure, and the incorporation into cognitive structure of the new concept, differentiated from previously learned, related concepts, are very important component processes in concept formation; in these respects his definition goes beyond Bruner's.

To summarise then, concept learning is considered to involve the identification of relevant attributes (which are themselves concepts already possessed, or relatable to specific concepts already possessed) and the rule combining these attributes. The important processes involved in concept learning are generally seen to be the formulation and testing of hypotheses regarding attributes, and the resulting selection of a set of predictive defining attributes. These processes are implied by the terms 'concept formation' and 'concept attainment.'

When we consider specific characteristics of science concepts, and the behaviour exhibited in learning science, the formulation of concept learning outlined above does not seem altogether appropriate. Much of the work described above considered adult learning, but this restriction would not present a major limitation on its use within Science Education, as much learning of science concepts occurs at a mature or near mature level of development. The differences that will be suggested below seem to be of a more fundamental and serious nature.

First, science courses frequently require the learning of <u>new</u> concepts, rather than the learning of a new grouping or hierarchy of already established concepts. In all except the most junior classes, pupils will generally be given a formal definition of the concept. If it is of the class or categorisation type, the definition will specify the predictive defining attributes (e.g. "a carboxylic acid is an organic compound that contains a COOH group"). Where the concept is of the formal or abstract type, the definition will include an exposition of the "intrinsic attribute properties",[41] (e.g. "A force is that which causes a change in the state of motion of a body"). In either case, as the attributes are specified for the learner, his subsequent behaviour can hardly be described as a process of formulating and testing hypotheses regarding the attributes, and selecting a set of predictive defining attributes.

One important process that must occur following the definition of a new concept is the building up of what has been variously termed the "intention", the "cachet specifique" (Bruner, from Michotte),[42] or the "generic meaning" (Ausubel). Because of the interconnectedness of the conceptual frameworks of many science disciplines, many science concepts (including some of the most generally applicable and fundamental concepts within the physical sciences) have a special characteristic; their defining attributes or intrinsic attribute properties are themselves essentially new concepts to the learner. The "scientific" generic meaning of the defined concept cannot, therefore, be acquired by considering the generic meanings of the attributes, which have not been established either. Instead, it must be evolved by generalisation and abstraction from concrete empirical experiences, and from the instances or situations in which the concept is shown to be used in learning which takes place <u>after</u> the definition has been given.

The fact that generic meaning must be established by learning that occurs <u>subsequent</u> to a definition, rather than from learning that has occurred prior to the definition, constitutes an important difference between the learning of science concepts, and the type of concept learning discussed above.

Ausubel does discuss the learning of concepts for which the defining attributes have been given. He calls this process "concept assimilation", and considers it to be the characteristic means of acquiring new concepts for adolescents and adults. Ausubel describes this process[43] as one in which pupils "learn new conceptual meanings by being presented with the criterial attributes of concepts and by relating these attributes to relevant established ideas in their cognitive structures." It is certainly true that the learning of many science concepts involves this process - for example, 'speed' defined as "rate of change" of "distance" with "time". However, because of the requirement that the defining attributes be relatable to relevant established ideas (i.e. to relevant prior knowledge), the process of "concept assimilation" as defined by Ausubel cannot be used to describe the learning of the type of science concept discussed above, (for example, the learning of concepts such as potential energy, magnetic flux, or entropy).

Ausubel also discusses a type of meaningful learning that gives rise to what he calls 'combinatorial meanings'.[44] He considers that this type of learning can occur when new propositions or concepts can be "non-arbitrarily related to a broad background of generally relevant content in cognitive structure by virtue of their general congruence with such content as a whole." Although Ausubel has relaxed the requirement of relatability to specifically relevant ideas, this type of learning is still dependent on prior learning; also the examples Ausubel gives as characteristic of this type of learning are all of relationships - e.g. he mentions the relationships between mass and energy, and heat and volume. The learning of such relationships is certainly important in science, and such relationships have an important role in the learning of the related concepts, (as will be considered below); nevertheless, for the reasons given above, the acquisition of generic meaning of many science concepts (for example, energy itself) cannot be described as learning that gives rise to combinatorial meanings.

To summarize then, the terms "concept formation" and "concept attainment" cannot usefully be applied to the learning of science concepts where learners are given a formal specification of the defining attributes or intrinsic attribute properties. The learning of some science concepts could be described in terms of "concept assimilation", but there remains a body of important science concepts for which the acquisition of generic meaning must occur in learning that takes place subsequent to the formal definition. No formal exposition of the process involved in this type of learning seems to have been given in the literature.

In the investigations considered above, the acquisition of a concept is seen as a simple yes/no dichotomy. An experimental subject is given one particular type of task, and on the basis of his perfomance is judged to "have" or "not have" the concept. It does not seem possible to describe the acquisition of many science concepts in terms of such a dichotomy. For example, a learner may be able to write the equation of the form $PbCl_2 \rightarrow Pb^{2+} + 2Cl^-$, and use the mole relation expressed correctly at a time when he cannot write the equation $H_2SO_4 + 2NaOH \rightarrow Na_2SO_4 + 2H_2O$ and use the mole relation it expresses correctly (or vice versa). If a pupil is successful in some tasks, it is difficult to say that he "has not" acquired the concept; on the other hand, if he is also unsuccessful in other tasks, it is difficult to say he "has" the concept. It would seem more accurate to say that he has acquired the concept <u>to a certain extent</u>; that is, he has reached a level or stage in attainment which enables him to perform one type of task, but he has yet to attain the level required for a different task.

This is not meant to suggest that a learner can manage some tasks when he has .5 of the concept, but must wait till he has, say, .75 of it before he is successful in others.

In fact, we would suggest that the learning of many science concepts is not a matter of acquiring The Concept, and that for that reason, it is not pertinent to ask whether a learner has acquired a concept. Rather, we would suggest that it is an evolutionary process, occurring over a considerable period of time - often open-ended - in which a learner holds a series of versions or states of a concept, with later or more advanced states being richer and more powerful than earlier states; and that the state of a concept held at any one time would be an important factor in determining the type of task that could be performed successfully. Thus, success in a particular type of task would both depend upon, and be an indicator of, the current state of a concept. It follows that, instead of asking whether a learner has acquired "a concept", we should ask what state of the concept he has achieved.

Earlier in this section, we have argued that the acquisition, or building up of, generic meaning of many science concepts is an important element in learning subsequent to the giving of a formal definition. The converse of this is that much learning that occurs in science is related to, or results in, the building up of generic meaning of concepts.

Finally, we would suggest that these two processes - the acquiring of different states of a concept, and the building up of generic meaning - are intimately connected, if not identical. That is, the shift from one state of a concept to another is accompanied by a clear change in form or content (or both) of the generic meaning a learner associates with that concept. Even where the "scientific" generic meaning of a concept has been acquired by "concept assimilation", later learning may give rise to changes in that generic meaning. Thus it would seem that the holding of different states of a concept, characterised by differences in generic meaning, is very generally applicable to the learning of science concepts.

At this point, it should be noted that the transition from one state of a concept to another does not depend solely on developmental changes, although this is certainly an important relationship. (The research in this area has already been mentioned).[6] Ausubel clearly identified two types of change:[45] (i) developmental changes in acquiring concepts (changes from one age to another), and (ii) characteristic sequential changes occurring in the cognitive properties of a given concept from early to late stages in its acquisition within a particular age level. The latter type of change seems just as important as the former, although it does not seem to have been investigated to the same extent. Either type of change in form or content of generic meaning - or, as Ausubel says, in the cognitive properties of a given concept - will be associated with a change in the state of the concept. For example, one important change in the form of generic meaning would be a change in its degree of abstraction or generalisation. Such a change could occur as a result of developmental growth that increased a pupil's ability to form an abstract conceptual representation, or it could occur as a pupil met the concept in an increasing variety of situations.

In deciding upon a term that could be used to describe the holding of different states of a concept, it seemed important to avoid any confusion with the terms "concept attainment" and "concept formation", and also desirable to emphasise the importance of generic meaning in characterising differing states of a concept. For these reasons, we have chosen to use the term "levels of conceptual understanding."

To give some substance to this description of the learning of science concepts, we will outline some differences that could be associated with different levels of conceptual understanding. At a very low level a pupil might be able to do no more than give a rotely learned statement of the concept definition. After some time, he could perhaps give a substantive (i.e. in his own words) definition of the concept; this could be described as a shift from rote to more meaningful learning of the defining attributes of the concept. At a more advanced level, he could give a substantive description of the relation between one concept and other concepts (e.g. between 'temperature', 'mean kinetic energy' and 'heat'). He could perhaps state the characteristics that distinguished situations in which a concept should be used from those in which it should not be used. e.g. A pupil could perhaps explain why a mole relation should be used in a neutralisation problem, and why the same type of mole relation should not be used in a precipitation problem. At a very advanced level, a pupil could perhaps give a substantive expression of the equivalence of alternative definitions of a concept (such as entropy), or explain why scientists had chosen to define a particular concept, or define it in a particular way. As Fensham[12] has pointed out, this level of conceptual understanding is important for science concept learning.

The examples of changes given so far could be described as occurring along an individual learner's dimension - that is, they represent an idiosyncratic response to presented material. However, changes may also be "forced upon" a learner. In moving to more advanced levels of study, <u>all</u> learners may be expected to reformulate the generic meaning of a concept, in preparation for a set of circumstances in which the "old" version of the concept is inadequate. For example, a new set of defining attributes of a concept may be given. Kempa and Hodgson[46] have considered a particular instance of this, namely the modification of learners' perceptions of a concept as they are given sets of increasingly abstract defining attributes of a concept. They sought to determine third, fourth and fifth year Chemistry pupils' "levels of acquisition" of each of several chemistry concepts by asking them to select from four given sets of defining attributes of each concept, that one that most nearly corresponded to their idea of the meaning of that concept. Each of these sets varied from "completely concrete" (e.g. "an acid is a substance like sulphuric acid. It is dangerous and can easily cause burns"), through two intermediate expressions to "abstract" (e.g. "an acid is a substance which can donate protons to another substance"). (Their term "levels of acquisition," while related closely to attribute perception, would seem to be similar to the term "states of a concept" employed here). One of their most interesting findings was that although fifth form pupils had been given the most abstract definition of each concept a considerable time (generally about a year) before the test was administered, only 25% of fifth year pupils on average selected the most abstract definition (although there was a clear preference for the two more abstract, rather than the two more concrete, definitions). In fact, they reported a clear relationship between the length of this 'maturation period' for a concept, and the preference for its most abstract expressions.

We would expect this time lag between the giving of a 'new' definition and the emergence of a new related level of conceptual understanding. We have already suggested that the generic meaning of a concept must be built up in learning that occurs subsequent to the giving of a definition. When a 'new' definiton has been given (whether the new definition differs from the old in terms of abstraction, or in any other way), it is very reasonable to expect that, for some time, a pupil's level of conceptual understanding will be determined by the generic meaning he has built on the foundation of an "older" definition, and that new generic meaning will emerge only gradually, following new learning, and will supersede the older generic meaning only when the pupil perceives the inadequacy of his former level of conceptual understanding.

Kempa and Hodgson also found no significant difference between the response patterns of "high IQ" and "low IQ" groups within each year sampled, a result which was contrary to their expectation. We have suggested that the degree of abstraction of generic meaning will depend on an individual's ability to form an abstract conceptual representation (i.e. his developmental stage) <u>and</u> on the situations in which the concept is met. In other words, the ability to form an abstract generic meaning can be a necessary condition, but is not a sufficient condition, for attaining a certain level of conceptual understanding. Although a learner's IQ may determine his <u>ability</u> to reformulate generic meaning, the number of "new" situations in which the "new" concept <u>must</u> be used (which will be related to the maturation period) will be very important in determining whether he <u>has to</u> reformulate it. This situation dependency, which necessitates the time lag between the giving of a "new" definition and the emergence of new conceptual understanding, also means that there will not necessarily be a relationship between developmental stage (or IQ) and level of conceptual understanding.

The need to reformulate generic meanings may also be imposed on learners by the introduction of new concepts or principles. For example, the introduction of the Heisenberg uncertainty principle will mean that learners must seek to attain a new level of conceptual understanding of say, momentum, if they are to be successful in a new series of tasks.

The position we have reached can perhaps be illustrated most succinctly by saying that we would not expect a fourth year pupil to have reached the same level of conceptual understanding of functional groups as a graduate university student. It would not be sensible to ask if both had acquired "the concept" of a functional group; the important question would be whether each learner had acquired a level of conceptual understanding appropriate for the tasks expected of him.

If we consider a particular concept, it may be possible to define a certain level of conceptual understanding in terms of a specified generic meaning, and in addition to identify a specific task which can be used to determine which pupils have attained that level of conceptual understanding. In the next section, two levels of conceptual understanding of functional groups will be defined and their associated criterial tasks specified.

2.5 The Experimental Questions and the Experimental Design

Before stating the experimental design proposed, the two hypotheses to be tested, namely

(i) that the difficulties were visual in origin,

(ii) that the difficulties were conceptual in origin

will be considered separately.

2.51 The Visual Difficulties Hypothesis

A test of the visual difficulties hypothesis essentially required the definition of a task that would determine whether or not pupils were sufficiently confused by the visual pattern aspects of formulae to inhibit their acquisition of the chemical content of the formulae.

Because it seemed impossible to give, in advance, an operational definition of 'sufficiently confused,' the following scheme was proposed.

First, pupils would be given an immediate recall task, in which they would be required to reproduce, one at a time, each of a series of previously shown patterns of known difficulty or complexity. The patterns to be used would have no chemical content, but would reproduce the pattern characteristics of extended structural formulae. A pupil's performance in this test, called the Pattern Test, would be described in terms of his 'Visual Score,' a variable that would indicate the complexity of pattern he had reproduced correctly.

It was assumed that the difficulty of patterns that could be memorized and reproduced correctly in a recall task would not cause confusion in ordinary class-room work. Therefore, a comparison would be made of the mean Visual Score (indicating the difficulty of patterns that could be reproduced on average) and the complexity of the sort of formulae used in C.H.E. equations. If the formula complexity were less than the mean Visual Score, the visual difficulties hypothesis would be contra-indicated. The opposite relationship would give tentative support for the visual difficulties hypothesis. To allow for this second outcome, an examination would be made of incorrect responses. If such responses were characteristically incomplete, but fairly accurate representations of the original patterns, it would seem that the mean Visual Score measured simply indicated the size of pattern or amount of information that could be memorized and recalled under the particular experimental conditions. If, however, incorrect answers appeared as confused or jumbled representations of the original patterns, we would have obtained evidence in favour of the visual difficulties hypothesis. In examining the incorrect responses, the reproduction of "forward" and "backward" representations of repeated groups in a pattern would be particularly noted, and an attempt made to identify any specific pattern characteristics that seemed to cause confusion.

It was decided to investigate the relationship between visual ability and performance in Chemistry in two ways. First, the mean Visual Score would be computed separately for fourth, fifth and sixth year pupils. It was known that sixth year pupils performed better than younger pupils in some tasks at least; thus, if the visual difficulties hypothesis were true, we would expect to observe a statistically viable and practically significant difference in mean Visual Score across years. Secondly, the correlation between pupils' Visual Scores and their achievement in Chemistry would be determined within each year.

2.52 <u>The Conceptual Difficulties Hypothesis</u>

Following their study of C.H.E. reactions, pupils are required to perform tasks such as writing or identifying equations illustrating a particular type of reaction, or determining the reactants or products related to a particular reaction; these tasks in turn necessitate the identification, or writing, of formulae for examples of a particular family. Several levels of conceptual understanding of functional groups could be useful in such tasks, but the observations reported in Section 2.3 suggested that we should test, at least initially, for low levels of conceptual understanding.

We defined two low levels of conceptual understanding that would be useful in performing the required tasks. The first of these would be characterised by a cognitive representation of the functional groups of the families studied - the $-C(=O)-O-H$, -O-H, and >C=O groups, and also the CH_3 and CH_2 groups, as chemical entities (that is, as chemically meaningful bricks from which molecules were built.) The behavioural task chosen to determine which pupils had reached this level of conceptual understanding was the recognition of the specified groups as units (e.g. physical units, not named units), when they were seen in an extended structural formula.

'Recognition' is used here in a technical sense, which can best be explained by analogy. If we meet a friend, we recognize him - that is, we see him and know him (even if we cannot recall his name) without concious thought. However, if we are asked to meet a stranger, we would be given a description of him, and would attempt to identify him by "ticking off" his described characteristics. Thus, recognizing a functional group as a unit involves instinctively noting it as one thing, and it is to be contrasted with identifying a particular collection of chemical symbols as a specific functional group, by "ticking off" the individual symbols. In the same way, an adult recognizes a word as a unit, whereas a young child sees it as a collection of letters that must be put together to make a word. Just as we recognize a friend even in unfamiliar surroundings, so we expected that pupils who had reached this level of conceptual understanding would recognize a functional group as a unit in whatever orientation it was drawn. It should be noted that the criterion of recognition could be used because pupils were not taught to recognize groups as units, so that such behaviour could validly be attributed to a pupil's conceptualisation of functional groups.

A second, and slightly higher level of conceptual understanding would be characterised by the cognitive representation of the functional groups >C=O, -C(=O)-O-H, and -O-H, as very important chemical entities.

Pupils who had achieved this level of conceptual understanding would recognize the five groups specified earlier as units, and in addition, would note the functional group(s) in a formula first, and note them correctly, whatever other details of the formula were not noted. As pupils were not taught explicitly to note the functional group(s) first, such behaviour could also be validly attributed to pupils' conceptualisation.

To determine which pupils had achieved these two levels of conceptual understanding, pupils would be given a second immediate recall task, using extended structural formulae instead of non-chemical patterns; the complexity of formula a pupil could reproduce correctly would be indicated by his 'Molecule Score.' The difference between a pupil's Visual and Molecule Scores would be used to determine whether or not he recognized the specified groups as units. (The method proposed for doing this will be described in detail in the next chapter.) It did not seem possible to determine with certainty whether or not pupils noted functional groups <u>first</u> from the results of the proposed Molecule Test. However, it would be possible to identify those pupils who characteristically reproduced functional groups correctly even though other details of a formula were not noted correctly, by an examination of incorrect formulae responses.

It was decided that the relationship between recognition of groups as units and performance in chemistry would be investigated in the same way as Visual ability and performance - that is, by making an across years comparison of a measure of functional group recognition, and by determining the correlation between such recognition and achievement in chemistry within each year.

It was realized that care would have to be taken in specifying the interpretation to be placed upon different outcomes of the formula recall test. It seemed valid to state that a result indicating that the majority of pupils at least recognized the five groups as units would contra-indicate the conceptual difficulties hypothesis (at least at the specified level of conceptual understanding), and that evidence indicating that pupils, additionally, characteristically reproduced functional groups correctly in responses that were incorrect overall, would provide even stronger contra-indication. However, a result indicating that few pupils recognized the groups as units, and reproduced the functional groups correctly in incorrect responses, could not be related to the validity or non-validity of the conceptual difficulties hypothesis without careful consideration of the precise results of the Pattern Test. A young child may have a stable cognitive representation of, say, a house; however, if he is not sufficiently skilled at reading, he may still not recognize the word "house" as a unit. In the same way, we felt that if the results of the Pattern Test indicated that pattern aspects of formulae were causing confusion in some way, the non-recognition of groups as units would not necessarily support the conceptual difficulties hypothesis.

While the fact that pupils had not shown any awareness of confusion due to the pattern characteristics of formulae justified the proposal of the recognition task as a test of the level of conceptual understanding, it was felt essential to include in the experimental design a pre-test of the combined recall tests, and to allow for the possibility that the pre-test results could necessitate a reformulation of the combined test procedure.

An objective test of a relationship between visual ability and recognition of groups as units could be obtained by computing the correlation between Visual Score and a measure of unit recognition within each year. The sample size to be used in the pre-test was determined by considering the number of pupils required to give a power of at least .8 (at α = .05) for observing a non-zero correlation coefficient if a moderate degree of correlation existed. The power of a test is the probability of observing a difference, significant at the specified level, in the sample results, where a specified difference exists in the parent population. In the present case, we decided that a population correlation coefficient of .5 would represent a relationship of practical consequence between visual ability and recognition of groups as units. Therefore, we required a sample size that would give a probability of .8 for rejecting the coefficient's being equal to zero, if, in fact, the population coefficient were at least .5. From Cohen's tables,[47] the required number was 30.

If an examination of the responses given in the pretest, and the correlation between visual ability and recognition, validated the use of the Combined Tests as a practical test of the two hypotheses, the Combined Tests would be given to a large representative sample of fourth, fifth and sixth year pupils.

In order to relate experimental results validly to the learning of Chemistry in Scotland as a whole, the sample to be used in a major experiment would have to be drawn from a representative set of Secondary Schools. It was decided to ask for the co-operation of 30 schools chosen to be representative in terms of location, pupil intake, sex, and private/state management. (This number allowed for an anticipated 30% failure of response.) The requirement that the samples of pupils used be representative of all Scottish pupils necessarily meant that the numbers of pupils involved would exceed that needed for a high power (.8 to .9) in the statistical tests proposed. For this reason, the number of pupils needed to achieve a particular power for each test proposed for the major experiment was not specified, (although the power of each test used will be reported with the results.) It was also decided that the results of the main experiment should be related directly to pupils' performance in tasks such as identifying family members. As the precise way in which this would be done could not be determined until the results of the main experiment had been obtained and analysed, the Experimental Design given below had to remain open-ended.

2.53 The Experimental Design

Five stages were proposed, namely:

1. Design and test a series of non-chemical patterns and a series of formulae.
   - (i) Design and test a system for assigning a 'Difficulty Number' to a pattern or formula that would indicate its complexity.
   - (ii) Design and test a system for assigning a 'Visual Score' and a 'Molecule Score' to each experimental subject, which would indicate the complexity of pattern and formula (respectively) he could reproduce correctly.
   - (iii) Define a variable that would measure a subject's recognition of the groups $-CH_3$, $>CH_2$, $-C(=O)-O-H$, $>C=O$, and $-O-H$, as units, and determine a method for validating its use.

2. Pretest the combined Pattern Test and Molecule Test with a sample of 30 subjects.
   - (i) Administer the two tests.
   - (ii) Examine the incorrect pattern and formulae responses.
   - (iii) Compute the correlation coefficient for Visual Score and measure of recognition.

(iv) Compute the mean Visual Score.

Contingent upon the outcome of the pre-test,

Either

3. Modify or reformulate the test procedure,

Or

4. Administer the Pattern and Molecule Tests to a large representative sample of Scottish Secondary pupils, who had just completed a fourth, fifth or sixth year Chemistry course.

   (i) Compute the mean Visual Score and the Mean of the recognition variable for each year.

   (ii) Compute appropriate correlation coefficients for

   Visual Score - Achievement in Chemistry (within each year)

   Recognition - Visual Score (within each year)

   Recognition - Achievement in Chemistry (within each year)

   (iii) Validate statistically the use of the variable defined to measure recognition.

   (iv) Determine the statistical significance of differences between mean scores of successive years.

   (v) Determine the significance of the correlation coefficients.

   (vi) Examine incorrect pattern and formula responses.

   (vii) Relate the results to the validity or non-validity of the two hypotheses.

5. Design additional experiments to relate the results obtained in the main experiment to performance in commonly required chemistry tasks.

CHAPTER 3

The Development and Validation of the Test Procedures

In this chapter, a detailed description will be given of the Pattern and Molecule Tests. Section 3.1 will be concerned with the construction of the Pattern Test. The Difficulty Number system and the Visual Score variable will be described, and their validation reported.

The Molecule Test and the corresponding Molecule Score will be described in Section 3.2. The technique by which the Visual and Molecule Scores were used to define a variable to measure the recognition of the specified groups will also be discussed.

Section 3.3. considers briefly the connection between Short Term Memory and the Combined Test.

A pre-test of the combined Pattern and Molecule Tests was administered to a group of 33 fifth year pupils in March, 1973. The results of the pre-test will be discussed in Section 3.4.

Overall, then, the chapter reports the implementation of Stages 1 and 2 of the Experimental Design. Much of this work has been reported elsewhere.[48]

3.1 The Pattern Test

The patterns to be used in the Test had to fulfil two requirements; they had to reproduce as closely as possible the pattern characteristics commonly found in extended structural formulae, but they had also to be 'non-chemical'. Ideally, we required that the only difference between the patterns used in the first test, and the formulae used in the second test, should be the chemical content of the latter. It seemed possible to achieve that situation by generating patterns from actual extended structural formulae. Two processes were used; in some cases, each chemical symbol was replaced by a dot, while in others, each chemical group was replaced by a simple geometric shape. The way in which such patterns preserved the structure of a formula can be seen by comparing patterns (i) and (ii) in Figure 3.1 with their parent formula.

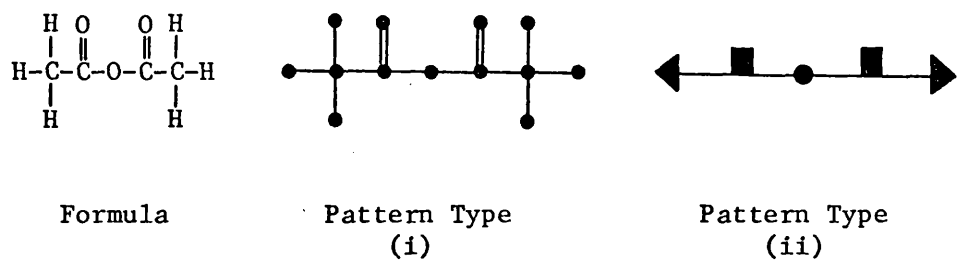


Figure 3.1 The patterns derived from a formula

The second stage in preparing the Pattern Test involved the construction of a system for specifying the difficulty or complexity of any given pattern. To do this as objectively as possible, a set of rules was drawn up which could be used to assign a 'Difficulty Number' to any pattern. The set of rules used is described in detail in Appendix 3.1. In devising the rules, we took the view that a pattern consisted of a number of equally important components (e.g. dots, shapes, double bonds, side branches, etc.), and that the greater the number of components the greater the difficulty of the pattern, with the proviso that the difficulty would also depend on the amount of symmetry and repetition within the pattern (for example, a line containing 5 dots would be simpler than a line containing 5 different shapes.) The rules specified the procedure for taking account of repetition and symmetry in assigning a score to each component of a pattern; the Difficulty

Number of the pattern was the total of these scores.

The Difficulty Number system thus allowed patterns to be ranked in order of difficulty, and also enabled patterns of equal difficulty to be identified. The latter property was an essential requirement for the Pattern Test. As mentioned in Section 2.5, the test was to be used to determine the complexity of pattern a student could reproduce accurately. For this reason, the test had to contain several examples of each Difficulty Number used, so that a pupil's performance could be assessed reliably at each Difficulty level.

The Difficulty Number system assigned either integer or (integer + 1/2) Difficulty Numbers to patterns. It was felt that Difficulty levels varying by only a 1/2 unit would represent an unwarrantedly fine division, and therefore it was decided to aggregate integer and (integer + 1/2) patterns into a Difficulty Group. So, for example, the Difficulty 7 Group could be composed of Difficulty 7, or 7 1/2, patterns. The Pattern Test was thus to consist of a number of Difficulty Groups, each of which would contain several patterns. Although the format of the Test could be prescribed to this extent, other important characteristics could be determined only by experiment.

First, the necessary range of Difficulty Groups had to be identified. Clearly, the first Difficulty Group would have to be within the ability of all students, and the highest Difficulty Group at the limit of students' ability, if each student's performance were to be assessed accurately.

Secondly, it was expected that the task of carefully observing and reproducing a series of patterns would require a high level of concentration. It was therefore necessary to determine the maximum number of patterns that could be shown before lapses in concentration would adversely affect, and therefore invalidate, the results. It was only when the necessary number of Difficulty Groups had been determined, and the maximum number of patterns was known, that the number of examples to be used within each Group could be fixed.

The time interval during which pupils could view a pattern (the Exposure Time), had to be long enough to enable pupils to observe the entire pattern, but not so long that they could rehearse it, and possibly devise a mnemonic or coding device for any confusing sections. A recording time (the time allowed for drawing the pattern just memorised) that just allowed pupils to reproduce all that they had memorised had also to be decided.

Various informal trials had suggested that a 10 second exposure time was sufficient for observing even the most complex patterns, and that after 20 seconds no further components were drawn. On the basis of similar immediate recall tasks,[49,50] it has been suggested that very little information can be rehearsed and stored in long term memory in a 10 second Exposure Time; that is to say, the pattern must be stored in, and reproduced from, Short Term Memory. The precise relevance of Short Term Memory will be considered later, but in general terms, the use of this Exposure Time would largely force pupils to record their 'first-off' perception of a pattern.

To avoid any emphasis on grouping within a pattern, it had been decided to draft patterns onto a rectangular grid. The final characteristic of the Pattern Test which had to be decided was the grid spacing (i.e. the symbol-symbol distance) which had to be large enough to allow all components to be perceived clearly, but not so large as to produce an unwieldy pattern overall.

A Trial Pattern Test was constructed so that appropriate values for these characteristics could be determined. The administration of this Trial Pattern Test also enabled a test of the validity of the Difficulty Number system to be made. If the difficulty numbers were a valid measure of the complexity of a pattern, one would expect to find the same number of correct responses (within experimental error) for patterns of equal difficulty, and, more importantly, that the number of correct responses would be a monotonically decreasing function of Difficulty. Failure to observe such a relation would lead one to question the validity of the Difficulty Number System.

3.11 The Trial Pattern Test

To maximize the probability of capturing the necessary Difficulty range within the Trial Test, consecutive Groups from Difficulty 2 (trivial patterns) to Difficulty 19 (very complex patterns) were constructed. Four patterns were chosen for each Group. These 72 patterns were then randomly assigned to positions within a viewing sequence, so that subjects would tend to make a more equal effort in observing each pattern. (If the patterns were shown in order of difficulty, subjects could become discouraged, and therefore not display their true ability, when the more complex patterns were shown.)

A 1 cm. grid spacing was used in drafting the patterns, which were reproduced as black-on-clear overhead transparencies. An Exposure Time of 10 seconds, and a Recording Time of 20 seconds, were used for the trial. The suitability of these values was determined by observations of the subjects during the experiment, and also by obtaining their opinions after the trial.

The Trial Pattern Test was administered to 40 first year Chemistry students at Glasgow University in February, 1973. An independent observer, who took no part in the administration of the test, assisted in the experiment, so that subjects' behaviour could be monitored continuously.

Each student was given an answer sheet, ruled to give a numbered sequence of squares; students were asked to record patterns in consecutive squares. Students were told that they would be shown each pattern for exactly 10 seconds, but that if the proposed 20 second recording period proved inadequate, it would be increased. They were also told that at the end of a recording period, the experimenter would say 'Next Pattern,' which would be their cue to return their attention to the viewing screen. Finally, to ensure that all subjects could observe the projected images with ease, and to illustrate the test procedure, an example pattern was shown for 10 seconds.

During the administration of the test, it became clear that the concentration span for this type of task was very short. In fact, the experiment had to be discontinued after showing only 48 patterns, and subsequent comments indicated that 40-44 patterns would have been more acceptable. Bearing in mind that subjects in the target population would be younger and less sophisticated than the first year University students, it was decided that the total number of patterns should be restricted, if possible, to about 35, but that it should on no account exceed 40.

Observations of subjects during the test, and their later comments, indicated that the Exposure and Recording Times were appropriate, and that the overall size and spacing of the pattern images were acceptable under experimental conditions. The tested values were therefore retained in later versions of the test.

The truncation of the test had one unfortunate consequence. Because of the random ordering of the patterns in the viewing sequence, the obtained responses were unevenly distributed across Difficulty Groups. Sufficient results were available to indicate that the critical difficulty region ranged from Difficulty 5 (the lowest Group in which incorrect responses occurred), to Difficulty 12 or 13 (in which the correct responses were effectively zero.) Given a required range of 5 to 12 or 13, and a desirable total of about 35 patterns, two possible test formats suggested themselves. The Difficulty Groups 5-12 could be used with four examples per Group, giving 32 patterns; alternatively, the groups 5-13 could be used, with only 3 examples given for the two highest Groups, giving a total of 34 patterns. The latter format was adopted, as it allowed more information to be gathered for only a slight increase in required concentration.

Although the correct response rate decreased monotonically across the Difficulty range 5 to 13, the results did not constitute a stringent test of the validity of the Difficulty Number system, because of the sporadic disposition of the obtained responses. However, this initial test provided sufficient information to enable the construction of a Pattern Test that could be used reliably to examine the validity of the system proposed for assigning the Visual Score, the second variable that had to be defined in connection with the Pattern Test. In addition, the administration of this second Pattern Test provided a further opportunity to test the validity of the Difficulty Number system.

3.12 The Visual Score

In essence, a variable that would indicate the complexity of pattern a pupil could reproduce correctly had to be related to the Difficulty Groups a student could handle, rather than indicating the total number of correct responses he made. A crude measure of this ability could have been obtained by taking as a pupil's Visual Score the number of the highest Difficulty Group he reproduced correctly. However, such a measure would have been deficient in that it took no account of a pupil's behaviour in responding to items of lower Difficulty, or to items of slightly higher Difficulty than the critical Group. To take account of such behaviour, a set of rules was drawn up for assigning a Visual Score to each pupil. These rules are reproduced in Appendix 3.2; briefly, they awarded a score on the basis of a pupil's performance in responding to consecutive Groups. In this way, a Visual Score of 7, for example, would indicate facility in reproducing patterns of Difficulty 7 or less, and little facility in reproducing more complex patterns. A Visual Score of 7.5 would indicate that a pupil could, in addition, reproduce one or two patterns of slightly higher Difficulty.

Whereas the validity of the Difficulty Number system required a particular relationship between numbers of correct responses for items of increasing Difficulty, the validity of the Visual Score system was most directly relatable to the sequences of responses of individual pupils. The set of rules defined would represent a valid measure of Visual Ability only if sequences of responses were characteristically consistent across consecutive Difficulty Groups.

3.13 Validation of the Visual Score System

The second, 34 item, Pattern Test was administered to a group of 19 fifth year and 14 sixth year pupils at a Scottish Secondary School. This group of 33 pupils represented a typical sample drawn from the target population. The conclusions drawn from this experiment could therefore be generalised within the target population.

The experimental procedure used for the Trial Pattern Test was repeated. Pupils' responses were therefore recorded in order of the viewing sequence. To test the validity of the Visual Score system, each pattern response was coded '1' for correct, or '0' for incorrect. The response codes for each pupil were then reordered to recover the original order of difficulty sequence.

These sequences were consistent in that each could be divided into three regions - an initial series of 'correct' Groups, then a borderline region in which perhaps one or two examples per Group were correct, and finally a series of Groups for which all responses were incorrect. Although the relative sizes of these regions differed from one individual to another, their pattern of responses was common for the sample. These results indicated that the Visual Score system, as defined, could provide a valid measure of the complexity of pattern that could be recalled and reproduced correctly.

The mean number of correct responses for patterns within each Difficulty Group is shown in Fig. 3.2. (The error bars represent the standard error for each mean.) Overall, these results indicated that the Difficulty Groups 5-13 represented the critical region for this sample, in that the correct response rate fell from approximately 90% to effectively 0% over the given range. It can also be seen that in no case was a mean significantly less than any subsequent means. This observed relation between the correct response rate and Difficulty was just that required to establish the validity of the Difficulty Number system, as a measure of the complexity of a pattern.

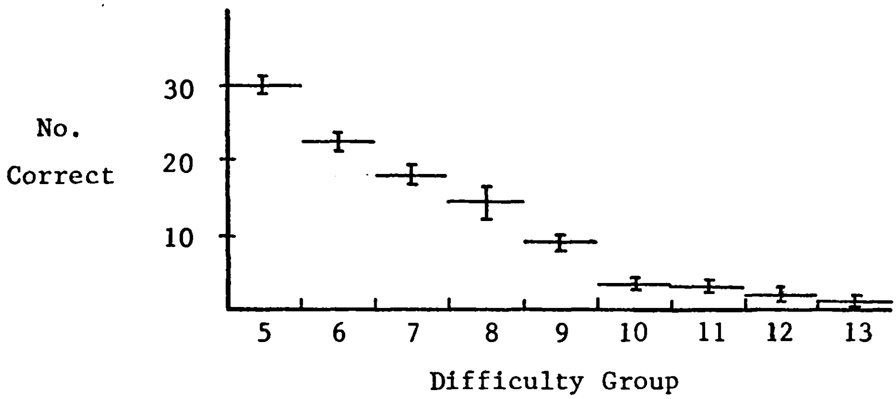


Figure 3.2 Mean number of correct responses per group

A detailed examination of the responses suggested that 5 patterns should be replaced. One pair of patterns, which differed only in the number of repetitions of a particular sub-group of components, had been included as an internal check of the Difficulty Number system. The responses for these two patterns suggested that some degree of memorization of the first had produced an anomolous response rate for the second pattern. In three other cases, there was some similarity between pairs of patterns, and although there was no indication that these similarities had affected the correct response rates, the safest procedure to adopt was the replacement of one pattern from each pair. The fifth pattern which was replaced had been generated from a formula containing a benzene ring. Comparison of fifth and sixth year pupils' responses for this pattern suggested strongly that the sixth year pupils recognized the skeletal benzene ring as a unit; this behaviour could not be expected from fifth (and even more certainly, fourth) year pupils. Obviously, the test could not contain a pattern in which the perceived components were different for different sections of the target population.

Pupils' behaviour during the experiment, and their later comments, indicated that the other characteristics of the Test, (the number of items, and Exposure and Recording Times) were appropriate. Overall, then, this experiment indicated that the Pattern Test, and its associated variables, could function as required within the target population.

3.2 The Molecule Test

The primary function of the Molecule Test was to determine whether or not a pupil recognized the groups already specified as units. To achieve this purpose, the Molecule Test was constructed on the assumption that pupils did not recognize any of the groups as units, with the truth or falsity of this assumption being determinable by an analysis of pupils' performance in the Test. In particular, it was assumed that extended structural formulae containing the relevant groups were patterns entirely equivalent to those used in the Pattern Test; the only difference between the two sets of patterns being that the components of which formulae were constructed included letters rather than dots or shapes. Based on this assumption, the construction of the Molecule Test simply involved the production of a second (formula) pattern test. That is to say, the Difficulty Number system was used to rate the complexity of formula patterns, and a series of Difficulty Groups containing 4 members per Group was obtained. One constraint operated in choosing formulae for the Molecule Test; no formula could be used that could be recognized 'in toto' (due to familiarity) by members of the target population. (The unfamiliarity of the formulae used in the Test was agreed to by a number of experienced Chemistry teachers.)

The administration of the Molecule Test replicated the administration of the Pattern Test, with one slight difference. The formula-patterns were drafted on a 1½ cm rectangular grid, as the 1 cm grid used for the Pattern Test resulted in a lack of clarity of certain letters used in formulae. The Visual Score system could be used to measure each pupil's performance on the Molecule Test, but, for clarity, the score assigned by the operation of that set of rules on the formula responses was termed the Molecule Score.

Under these conditions, we would expect to find no difference (within experimental error) between a pupil's Visual Score and his Molecule Score, if in fact he perceived formulae only as 'letter-and-line' patterns - that is, if he failed to recognize the specified groups as units. If, on the other hand, a pupil did recognize the specified groups as units, we would expect his Molecule Score to be greater than his Visual Score. The group $-\overset{O}{\overset{/\!/}{C}}$ O-H, for example, contributes three unique components ('C', 'H', '//') and one repeated component ('O') to a formula, and each of these components would contribute to the Difficulty Number under the assumption that the formulae were just patterns. However, if this collection of components were actually recognized as a unit - that is, as effectively one component - the complexity of the formula pattern 'as perceived' would be less than the rated complexity. Thus, the effective Difficulty Number would be less than the assigned Difficulty Number. If the effective Difficulty Number were not greater than the pupil's Visual Score, we would expect him to reproduce the formula correctly. In this way, a pupil whose Visual Score was, for example, 7 could perhaps reproduce formulae rated as Difficulty 8, 9 or 10, and hence obtain a Molecule Score of 10, thus obtaining an increment of 3.

Of course, it does not follow that any increment between Visual and Molecule Scores could be taken as evidence of recognition of the specified groups as units. (For example, a small increment could be expected if a pupil recognized just CH_2 as a unit.) Nor could one particular value of increment be used as a criterion of recognition for all pupils. Given that the assigned Difficulty range of the Molecule Test must be finite, a pupil whose Visual Score was 5 would clearly have the possibility of achieving a larger increment than a pupil whose Visual Score was, say, 11. The criterial increments were, in fact, those that would be obtained, given the criterial performance - the correct reproduction of those formulae whose effective Difficulty was less than or equal to the pupil's Visual Score. The criterial or 'expected' increments were thus different for each value of Visual Score. The way in which these Expected Increments were calculated is described in Appendix 3.3.

Following the administration of the combined Pattern and Molecule Tests, the series of actions proposed for determining whether or not the specified groups were recognized as units was:

(i) Determine the pupil's Visual Score

(ii) Determine his Molecule Score

(iii) From these, compute his Actual Increment

(iv) Select the Expected Increment corresponding to his Visual Score

(v) Evaluate the Ratio (Actual Increment/Expected Increment).

The criterion for the recognition of the specified groups as units was thus a value for the Ratio greater than or equal to 1. The Ratio itself was taken as the variable that would measure a pupil's recognition of the groups as units.

The validation of this system required that each pupil's Actual Increment should be shown to have been obtained by reproducing correctly those formulae whose effective Difficulty was less than or equal to his Visual Score, when the specified groups were considered as single components. If Increments were generally obtained by reproducing correctly an arbitrary set of formulae, the ratio would not be a valid measure of the recognition of groups as units.

The only characteristic of the Molecule Test that was not predetermined by the tested characteristics of the Pattern Test was the range of Difficulty Groups. To allow for increments, the Molecule Test had to extend beyond the upper Difficulty limit of the Pattern Test, but, at the same time, the constraint of a maximum of 40 items still obtained. It was decided that the range Difficulty 6 to Difficulty 15 should be tested in the first instance; as in the Pattern Test, only three items were included in each of the last two Groups, giving a total of 38 items. As this was the only untested property, it was decided that a separate trial of the Molecule Test would not be made, but that the suitability of the range chosen would be determined during the administration of the combined Pattern and Molecule Tests in the Pre-test proposed as Stage 2 of the Experimental Design. (It should perhaps be noted that the validity of the Ratio variable could be tested only following an administration of the combined tests.)

Further understanding of the operation of the combined Pattern and Molecule Tests may be gained by considering the particular constraints imposed on immediate recall task performance by certain characteristics of Short Term Memory. Therefore, before considering the Pre-test, a brief description of Short Term Memory will be given, and its particular relevance to the combined tests will be discussed.

3.3 Short Term Memory and the Combined Tests

'Short Term' Memory or 'Immediate' Memory has been distinguished from Long Term Memory.[51] This distinction is inherently attractive because it corresponds to an intuitive awareness of two types of memory process; the 'short term' type, that enables the effortless recall of a small amount of information (such as a new telephone number) over a short period of time, and a second type, which may require a more conscious effort of memorization initially, but which is effective for even large amounts of information over a much longer interval of time. Such a distinction would also seem to be of practical value, in that several studies have reported results suggesting strongly that two different memory mechanisms can operate,[52,53,54] one effective over short retention times and the other less sensitive to retention time; these mechanisms may operate simultaneously.[55,56]

Two features of Short Term Memory are germane to the present discussion. First, Short Term Memory is associated with a limited capacity for the storage and subsequent retrieval of information.[57] Secondly, and most importantly, the limitation is on the number of 'chunks' of information that may be stored and retrieved. The term 'chunks' was first employed by Miller;[58] his suggestion that Short Term Memory capacity was about 7 $\pm$ 2 chunks has received wide acceptance. A chunk is not a fixed, observer independent quantity of information; on the contrary, it is simply what the observer perceives or recognizes as a unit. For instance, a word, a letter, or a digit could be a chunk, and the capacity of Short Term Memory is in each case about 7 chunks - that is, approximately 7 words, 7 letters, or 7 digits. The total amount of information (in terms of the fixed technical quantity, the 'bit' of information) which can be stored in Short Term Memory will thus depend on the amount of information contained within each of the 7 chunks. A clear demonstration of the relevance of chunks to Short Term Memory has been given in a study[59] which showed that there was little difference between the retention of a three letter series and a three word series, but that there was a significant difference between the retention of these 'three chunk' series and the retention of a single word.

The storage of information as chunks suggests that a pupil who had shown that he could memorize and reproduce a maximum of, say, 7 unrelated letters, could nevertheless be expected to reproduce the 9 letter sequence 'b c a t f n s l o' if he perceived it as the 7 chunk series 'b cat f n s l o.'

In an analagous way, a pupil whose Short Term capacity for patterns was represented by a Visual Score of 7, for example, could be expected to reproduce formulae of greater rated Difficulty if he perceived some of the components of each formula as a chunk (i.e. recognized a specific group as a unit.) The Expected Increments, defined in the last section, represent the increase in information that could be stored in Short Term Memory, given a particular, specified chunking procedure.

The inferring of a level of conceptual understanding from an observed chunking procedure is, in some ways, the inverse of the position adopted by De Groot. In a series of experiments,[60,61] he compared the ability of known chess masters and novices to reproduce the positions of chess pieces after a 5-10 second exposure time. He reported that the performance of the two categories of player was identical (about 6 positions recalled) when the pieces were positioned arbitrarily, but that when a game position was used, the novices were unable to improve their performance, whereas the masters could reproduce almost all the positions correctly. This behaviour was attributed to their ability, due to their superior mastery of chess, to 'chunk' the information presented to them.

In the present case, we would expect that 'chemistry masters' and 'chemistry novices' would show similar ability in reproducing non-chemical patterns, but that the two groups could be distinguished by virtue of their differing ability to chunk the information presented in structural formulae.

In Section 2.52 consideration was given to the possibility that pupils might not be able to recognize a group as a unit if severe confusion were engendered by the pattern characteristics of formulae. That is to say, the ability to chunk information presented in that particular format could depend on Visual ability as well as on the level of conceptual understanding. The ability to chunk could be taken as an indicator of the level of conceptual understanding only if no significant relationship existed between chunking and Visual ability. This underlines the necessity for a critical examination of the correlation between Visual Score and Ratio (the measure of group chunking) in the Pre-test.

3.4 The Pre-Test

The combined Tests were administered to two groups of fifth year pupils (N=33) at one Scottish Secondary School, in March, 1973. The experimental procedure was essentially that used for trialling the Pattern Test. Pupils were informed that the experiment was part of a study of learning in Chemistry, and that they would be required to observe patterns, then reproduce them from memory onto the numbered answer squares. The Exposure and Recording Times were stated, and pupils were told that a cue would be given to indicate the end of each recording time. Again, an exemplar pattern was shown before the first item of the Pattern Test. Pupils were given a fifteen minute rest period between the administration of the two tests, and it was only immediately prior to the start of the Molecule Test that pupils were told it would contain structural formulae. In this way, the chance of pupils' perceiving the patterns as formula analogues was minimized. Such a perception could enable pupils to employ a chunking procedure in the Pattern Test, which would invalidate the use of the Ratio as a measure of recognition of groups as units.

3.41 The Pattern Test Results

The mean number of correct responses per Group was again a monotonically decreasing function of Difficulty. The number of correct responses for items that were used in both the Pre-test (School 2) and the earlier school trial (School 1) are shown in Figure 3.3. The totalled responses show the same relation between number of correct responses and Difficulty. Thus, the results from each school individually, and the combined results, indicated that the Difficulty Number system was a valid measure of the complexity of a pattern. An examination of each pupil's sequence of correct responses again showed consistent behaviour across Difficulty Groups, giving a further demonstration of the validity of the Visual Score system.

The mean Visual Score of the 33 Pre-test subjects was 7.5 (S.D. = 1.3). An examination of pupils' responses showed a very interesting phenomenon - the way in which the responses had been written suggested very strongly that pupils had read and memorized each pattern from left to right, as though it were a word, noting first the long central chain and then the side branches and their positions. The most characteristic mistakes were the transposition of side branches, or

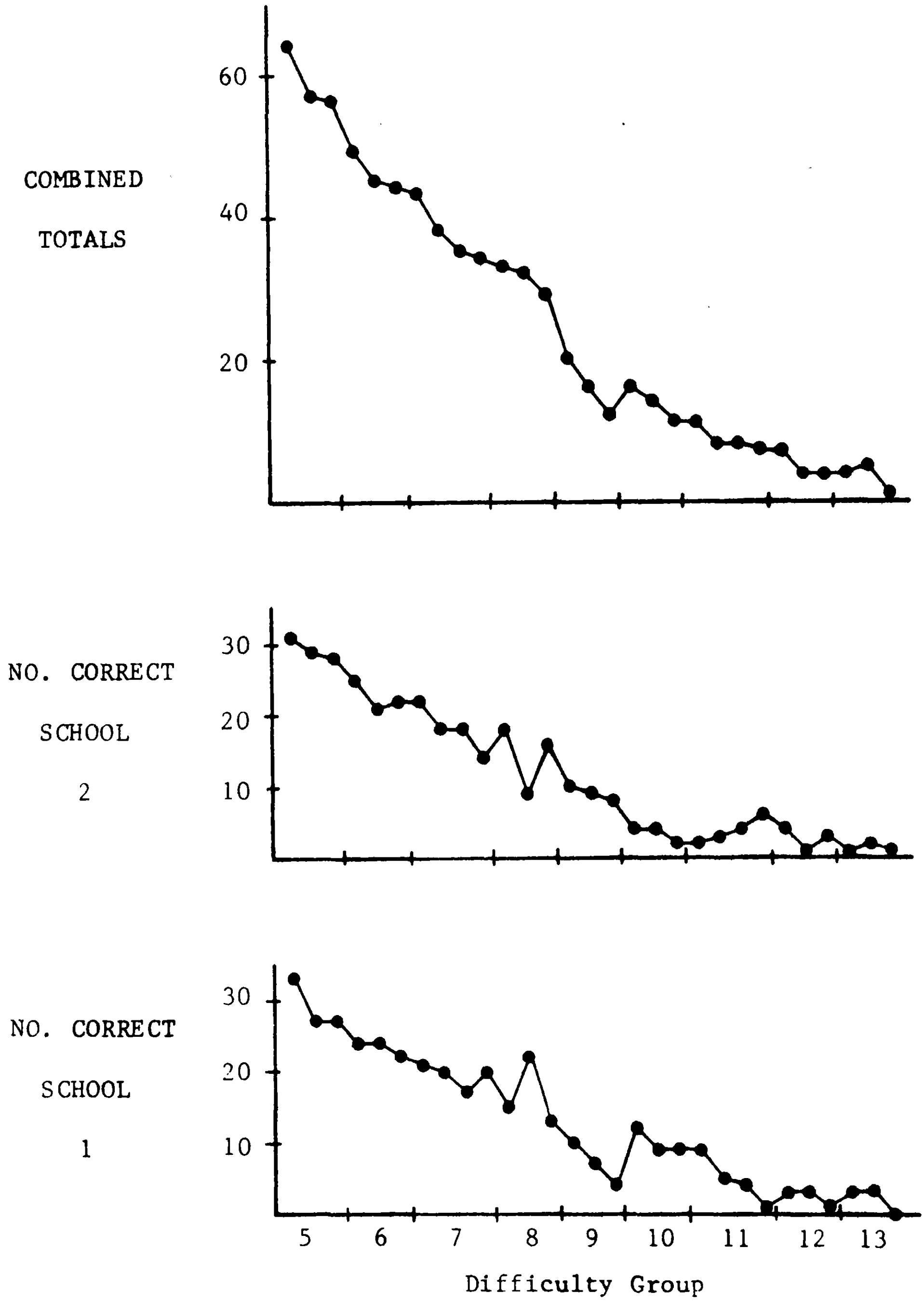


Fig. 3.3 Number of Correct Responses for Pattern Test Items

their omission, particularly toward the right hand end of a pattern. Overall, incorrect responses tended to be either incomplete, or inaccurate but sensible reproductions of the original patterns.

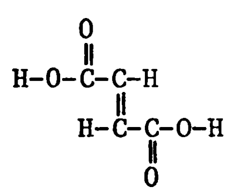
Three patterns, having a common characteristic, were to some extent, exceptions to the rule. Their common property was a side chain, of approximately the same length as the central horizontal chain, located at the left hand end of the pattern. The responses for patterns containing similar side chains at the middle or right hand end were in no way exceptional This fact, together with the actual representations of the three patterns, suggested that when pupils began reading such patterns, they could not decide whether to take the vertical or horizontal chain as a "base line." Commonly, side branches were transposed between the vertical and horizontal lines, and the intersection of these two lines was also often incorrectly reproduced. The components and groups of components did seem to have been perceived reasonably accurately, suggesting that there was not overall confusion. The effect of this characteristic could be described as having increased the difficulty of the patterns in a way not accounted for by the Difficulty Number system. It was decided to retain these patterns in the Test so that more information could be obtained about them, but to award each pattern an extra point for this characteristic.

The responses for patterns that contained a group of components and its mirror image showed no evidence of confusion. The Difficulty Number system assumed that the symmetrical reproduction of a group reduced the Difficulty of the pattern more than the simple repetition of the components of a group. Now, the number of correct responses for symmetrical patterns suggested that their assigned and perceived Difficulty were identical. This suggested that pupils had perceived the two sets of components as mirror images of the same group.

3.42 The Molecule Test Results

The lowest Molecule Score obtained was 7.0, and the highest 14.667, which indicated that the range of the Molecule Test (6-15) was adequate. Pupils reported that the number of items in the Molecule Test was not excessive. It therefore seemed that the characteristics of the Molecule Test were appropriate for the target population.

An examination of the responses suggested strongly that pupils 'read' the formulae as words, just as they had read the patterns. As with the patterns, incorrect responses tended to be incomplete, or inaccurate but sensible, reproductions of the original formulae. The only exception was the formula



for which many jumbled responses were given. Responses for formulae that contained a functional group and its mirror image showed no evidence of confusion.

Finally, there was only one instance in the thousand or so responses of a functional group's being reproduced correctly when the rest of the formula was completely wrong.

3.43 The Combined Test Results

The Pearson product-moment correlation coefficient calculated for Visual Score-Ratio was -0.1. The 95% confidence interval[62] was (+.08, -.27). In this work, a confidence interval will be quoted for each computed mean or correlation coefficient. These intervals are informative because it is impossible to reject any hypothesis, at the 5% level, that equates the value of the population parameter with any value lying within the stated interval. Therefore, the confidence interval represents the uncertainty in the population parameter after the measurement of a sample parameter.

Given the confidence interval about the Pre-test correlation coefficient, it was impossible to reject the hypothesis that the correlation between Visual ability and group recognition was zero. Furthermore, for a sample size of 33, the power to reject H_o: $\rho = 0$ in favour of H_1: $\rho = .5$ was .86. Therefore, in accepting the hypothesis that the correlation was zero (rather than at least 0.5) the probability of committing a type II error (i.e. of accepting H_o when in fact H_1 was true) was .14.[47] This result indicated that it was extremely unlikely that the ability to chunk the specified groups was related to Visual ability.

The Pre-test thus provided two important results:

1. There was no evidence of confusion shown in the responses to the Pattern Test (with one possible exception);
2. There was no evidence to suggest that the ability to chunk groups within formulae was related to Visual ability.

These two results validated the use of the combined tests in determining whether or not the specified groups were recognized as units.

3.44 The Augmentation of the Ratio Measurement

The mean value of the Ratio was 0.48 (95% Confidence interval = (.33, .61). Only two pupils had a ratio of 0.9 or greater. Assuming for the moment the validity of the Ratio as a measure of group recognition, this result would indicate that only these two possibly recognized the specified groups as units. The Combined Tests therefore provided very little information about the way in which the majority of pupils perceived formulae.

The technique described in Appendix 3.3 could be used to determine the increments to be expected if any specified set of groups were recognized as units. Therefore, in order to obtain a more detailed description of pupils' behaviour, two additional sets of groups were defined. The original set of five groups was re-termed the 'Class I' set, and the additional sets, the 'Class II' and 'Class III' sets. The three sets were, then:

CH_3, CH_2, OH, C = O, $\overset{O}{\overset{\|}{C}}$–O–H - Class I

CH_3, CH_2, OH, = O - Class II

CH_3, CH_2, OH - Class III

The component groups included in the two additional Classes were chosen because the formula responses suggested that they had been written as units. (For instance, the OH group was always drafted as -O-H, but was often reproduced as -OH.)

The Class II and Class III increments for each Visual Score value were determined by applying the technique described in Appendix 3.3.

The results of the Pre-test were then re-analysed, using the following procedure:

(i) The Class I, Class II and Class III increments corresponding to a pupil's Visual score were selected.

(ii) Three ratios, $\frac{\text{actual increment}}{\text{Class I increment}}$ (Class I ratio),

$\frac{\text{actual increment}}{\text{Class II increment}}$ (Class II ratio),

$\frac{\text{actual increment}}{\text{Class III increment}}$ (Class III ratio),

were computed.

(iii) The ratio closest to 1 was identified.

(iv) The pupil was assigned to Class I, Class II or Class III according to which of the calculated ratios fulfilled condition (iii), unless,

(v) His actual increment was zero (i.e. he recognized no groups as units.) He was then assigned to Class IV.

The categorization of pupils obtained by this procedure was: 2 pupils in Class I, 18 in Class II, 12 in Class III and I in Class IV.

It should perhaps be emphasised that, while this procedure (if shown to be valid) could provide more information about the perception of formulae, it was only the recognition of the original, or Class I, groups as units that could be related to the level of conceptual understanding of functional groups.

The requirement originally proposed for the validation of the Ratio as a measure of group recognition was that the actual increment of each pupil should have been achieved by reproducing correctly those formulae whose effective Difficulty was less than or equal to his Visual Score, when the specified groups (the Class I groups) were taken as units. The requirement for the validation of the procedure that assigned pupils to one of the four categories defined above became more complex.

It was possible to specify, for any Visual Score, three sets of formulae that should have been reproduced correctly corresponding to the recognition of Class I, Class II, or Class III groups as units. That is, for each value of Visual Score, it was possible to specify a Class I, a Class II and a Class III pattern of correct responses. We required that each pupil's actual pattern of correct responses matched the pattern of correct responses corresponding to his Visual Score and assigned Class better than the patterns corresponding to his Visual Score and any of the other Classes.

The sample size prevented a statistical test of this requirement. However, a visual comparison of each pupil's actual set of correct responses with the three sets of responses corresponding to his Visual Score was made. In each case, the best agreement was obtained between the actual set and the 'assigned Class' set of correct responses.

Overall then, the Pre-test results indicated that the Combined Tests could function as required. It was therefore decided to proceed to Stage 4 of the Experimental Design, the administration of the Combined Tests to a large, representative sample of fourth, fifth, and sixth year pupils.

(The reliability, or reproducibility of the Visual and Molecule Scores is discussed in Appendix 3.4).

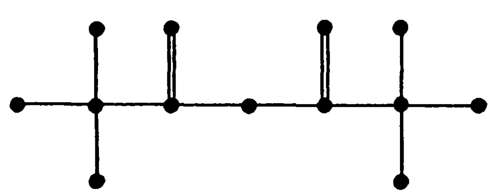
APPENDIX 3.1

The Difficulty Number System

In assigning a Difficulty Number to a pattern or formula, the following rules were applied:

1. One point was awarded for the first instance of each component (i.e. each dot, shape, double or triple bond, or chemical symbol.)
2. ½ point was awarded for each repetition of a particular component.
3. One point was awarded for a central line containing more than four components.
4. One point was awarded for a side chain containing more than one component. (Obviously, a side chain and a central line could intersect in a component; such a component was considered as belonging to the central line.)
5. Where a group of components was repeated, each repetition was awarded half the score assigned to the previous instance of the group.
6. One point was subtracted from the total score if a group was repeated as its mirror image.

Examples:



(a) This pattern contains a group repeated as its mirror image. 1 point is awarded for the central dot. (Each other dot is awarded ½ point.)

The score awarded to one group is:

(6 x ½) for dots; 1 for double bond; 1 for side chain;

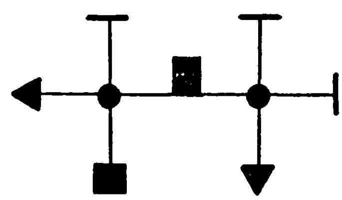
The total for the group is therefore 5.

The repeated group is awarded 2½.

The central line is awarded 1.

Total = 1 + 5 + 2½ + 1 = 9½.

1 point is subtracted for the mirror symmetry, giving a Difficulty Number of 8½.



(b) Considering the central line:

1 point is awarded for the triangle, one circle, the rectangle, the bar, and the central line itself. ½ point is awarded for the repeated circle. This gives 5½ points.

Considering the side chains:

1 point is awarded for the square, and ½ point for each of the two bars, and the triangle. 1 point is awarded for each side chain. This gives a further 4½ points.

The Difficulty Number is therefore 10.

(The complete set of patterns and formulae used have not been reproduced here because they are just one set of items that can be generated using the set of instructions given above. It is this generating set which is of fundamental importance).

APPENDIX 3.2

The Visual Score System

The experimental conditions required pupils to work under pressure, and to record responses quickly. Therefore, the system made an allowance for careless mistakes in the following way. A Group was considered 'correct' if at least three out of four (or two out of three) items were reproduced correctly, but only until a Group occurred in which more than one response was incorrect. Once this occurred, a Group was considered 'correct' only if all 4 responses were correct.

To determine a pupil's Visual Score, his responses (transposed to give the original order of difficulty) were examined Group by Group. The highest consecutive Group that met the 'correct' criterion was identified, and the pupil assigned that integer value. The fraction of each subsequent Group correct was added to this integer.

A careless mistake could be called an 'unlucky chance'. The system also allowed for a 'lucky chance'. Operationally, this was defined as a correct response occurring after two Groups that were completely incorrect. (These Groups were not necessarily consecutive). Thus, when two completely wrong groups had been encountered, no further Groups were scanned, and the totalled score became the pupil's Visual Score.

Where a pupil did not meet the 'correct' requirement for the first Difficulty Group (Group 5 for the Pattern Test, or Group 6 for the Molecule Test) it was assumed that he would have had an immediately prior Difficulty Group 'correct'. He was therefore awarded the default value of 4 (Pattern Test) or 5 (Molecule Test), and the appropriate fractions were then added to this integer.

Examples

Group	5	6	7	8	9	10	11	12	13
Student (a)	1110	1111	1101	1111	0000	0000	0000	100	000
Student (b)	1101	1100	1110	1001	1010	1000	0000	000	000
Student (c)	0011	1000	0000	0000	0000	0000	0000	000	000

The three sets of responses shown above are typical of the results obtained from the Pattern Test. A "1" represents a correct response, and "0" an incorrect response.

Student (a) meets the 'correct' requirement for Groups 5 to 8 inclusive, because he has no more than one incorrect response per Group until Group 9.
His score would be 8 + 0 + 0 = 8. (The correct response in Group 12 is not scored, because Groups 9 and 10 contain no correct responses).

Student (b) meets the 'correct' requirement for Group 5 only, as there are two incorrect responses in Group 6.
His score would be 5 + .5 (Group 6) + .75 (Group 7) + .5 (Group 8) + .5 (Group 9) + .25 (Group 10) = 7.5.

Student (c) has no 'correct' Group. He is therefore awarded the default value of 4.
His score would be 4 + .5 (Group 5) + .25 (Group 6) = 4.75.

These examples illustrate the weighting given to consistency of response by the system. Student (a) has only 1 more correct response than Student (b) (who also has correct responses for more complex items) but his greater consistency results in a score .5 greater than that awarded to Student (b). For Student (c), a score of 5 would have been awarded had the 3 correct responses occurred in Group 5.

APPENDIX 3.3

The Expected Increments

To calculate the Expected Increments between Visual Score and Molecule Score, given that each of a specified set of groups was recognized as a unit:

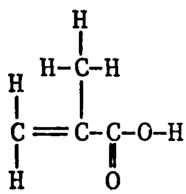
1. A geometric symbol was assigned to each group.
2. The groups were replaced in each formula by the appropriate symbols.
3. The Difficulty Number of each reduced formula-pattern was determined.
4. It was assumed that a pupil having Visual Score N should give correct responses to all reduced formula patterns whose Difficulty Number was less than, or equal to (N + 1/2). (N + 1/2) was taken as the critical value, because a Difficulty Group N contained both N and (N + 1/2) Difficulty Items. The number of correct responses expected within each Difficulty Group was determined.
5. The Visual Score System was then applied, to determine the Molecule Score that would be obtained given the particular set of correct responses identified in (4).
6. The Expected Increment was then obtained.

The process was repeated for each integer value N from 4 to 13. The increments expected for non-integer values of Visual Score were determined by interpolation. The Expected Increments for each Visual Score are listed below.

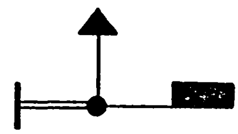
Visual Score	Expected Increments		
	I	II	III
4	3.75	2.25	2.0
5	4.75	3.0	1.5
6	5.5	3.75	2.25
7	4.75	4.0	3.0
8	4.66	3.5	2.25
9	5.33	4.33	3.5
10	5.0	5.0	5.0
11	4.0	4.0	4.0
12	3.0	3.0	3.0
13	2.0	2.0	2.0

As can be seen, the three increments for each of the Visual Scores 10 and above are identical. Unfortunately, the computational error which had indicated a difference between Class I and Class II increments up to Visual Score 12 was detected only after the combined Tests had been administered to the large sample selected. In the event, the correction of this error did not necessitate any change in the Pre-test classification of pupils, and affected only some 0.25% of the large sample. However, if this experiment were to be repeated, the formulae used in Groups 10 to 15 of the Molecule Test should be chosen so that a difference between at least Class I and Class II Expected Increments should obtain certainly up to Visual Score 12.

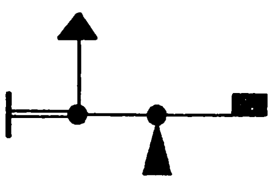
The set of reduced patterns associated with one of the formulae used is shown in Fig. 3.A3.1, together with the Difficulty Numbers.



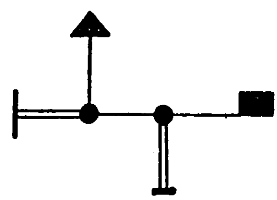
Formula (Class IV)
(13)



Class I pattern
(6)



Class II pattern
(8½)



Class III pattern
(10)

Fig. 3.A3.1 A formula and its reduced patterns

The four Class I-Class IV Difficulty numbers associated with each Molecule Item are shown in Table 4.A5.2 (Appendix 4.5).

APPENDIX 3.4

The Reliability of the Test Procedures

A conventional investigation of the reliability of a test procedure involves administering the test to the same subjects twice, and determining the correlation coefficient for the two sets of scores. Such a technique was not appropriate in the present case. If the Combined Tests had been given twice, with a period of a week or two between administrations, there would have been a very real possibility that behaviour in the second test would have been influenced by the experience of participating in the first test. In spite of this, a high correlation coefficient could have been obtained; this, however would not necessarily have demonstrated the reliability of the test procedures. Even when raw scores are used in computing a correlation coefficient, the coefficient provides a measure of the extent to which the rank ordering of pupils is constant across the two tests. In the present instance, maintaining the same rank order would not be a sufficient condition for the reliability of the procedures (i.e. the procedures used to determine the Visual and Molecule Scores).

We would require each pupil to achieve the same Scores in both administrations of the test (within experimental error). That is, we would require that the mean of the differences between the two sets of Visual Scores, and the mean of the differences between the two sets of Molecule Scores (in both cases, treated as dependent samples) be not significantly different from 0.

Clearly, it would not be valid to apply this test when participation in the first experiment could affect the outcome of the second. This problem could be overcome, presumably, by choosing a larger time interval between test and re-test - perhaps two to three months. In practice, this would have been difficult, given that most pupils within the target population were involved with public examinations very shortly after their participation in the Combined Tests. More importantly, however, the increase in age, and any learning that occurred during a long interval, could affect either the Visual ability or the level of conceptual understanding of pupils. Therefore, a comparison of scores could again not be used for determining the reliability of the Test systems. (It should perhaps be noted that the Visual Score system was not susceptible to a Split Halves[16] analysis; the possibility of an intrinsic difference between N and (N + 1/2) Difficulty patterns precluded the use of this technique in examining the reliability of the Difficulty Number system, because the number of integer patterns was not constant for all groups).

Because it was realized that the reliability of the procedures could not be tested directly, particular care was taken to make the Visual Score system robust as well as sensitive. Given the definition of 'correct' used in this system, it is clear that a pupil could obtain the same Visual Score by giving any one of a number of equivalent sets of responses. For instance, there are five ways of obtaining a 'consecutive' Group 'correct', viz. all correct, or any one of four incorrect.

The results obtained from the two schools do, however, provide indirect evidence for the reliability of the tests procedures, in that they could have provided evidence of unreliability of the procedures, but did not do so. First, it was noted (Sections 3.13 and 3.41) that the sequences of pattern responses were characteristically consistent across Difficulty Groups; consistency of responses within Groups is implicit in this behaviour. That is to say, pupils reacted in the same way to patterns of the same Difficulty, even though these patterns were randomly distributed through the viewing sequence. Had this behaviour not been observed, the Visual Score system would have been considered invalid, or, at best, unreliable.

In Section 3.44 it was noted that pupils characteristically responded in the same way to a set of formulae defined by Visual Score and assigned class. Again, this behaviour would not have been expected had either the Molecule Score or the Ratio been invalid or unreliable.

Finally, three of the four groups of school pupils to whom the Pattern Test was administered, consisted of fifth year pupils. The mean Visual Scores for these groups were: 7.4 (S.D. 1,3), 7.42 (S.D. 1.1), and 7.62 (S.D. 1.32). While it was in no way necessary that these means should be identical, gross differences between the means would have cast doubt on the reliability of the Visual Score system.

Thus, although the reliability of the procedures could not be positively demonstrated, the failure to find evidence of unreliability where such unreliability could have been expected to produce observable effects, justified the transition from Pre-testing to the administration of the Combined Tests to a large representative sample.

CHAPTER 4

The Combined Tests: Experimental Results and Analyses

The administration of the Combined Tests as a full scale experiment and the series of analyses outlined in Stage 4 of the Experimental Design (Section 2.5) provided a multiplicity of results to be considered in testing the Visual Difficulties and the Conceptual Difficulties Hypotheses. Because of their number, and because, in several instances, a result could not be applied to test the hypotheses without consideration of other results, the Experimental Data and the hypotheses will be discussed separately. First, the experimental results will be described, and the associated statistical analyses presented and discussed. The status of the two hypotheses will then be considered, in the final section of this Chapter.

The results, although varied, fell fairly naturally into three groups, which will be used to provide a structure for their presentation. The first group consisted of the results obtained from the Pattern Test alone. The "raw data" - the number of correct responses for each Test Item, and the nature of students' responses - will be described and discussed. Probably the most important set of results considered in this section is the set of mean Visual Scores for each year.

The raw data obtained from the Molecule Test, and the results obtained from the Visual and Molecule Scores taken together, formed the second group of results. A statistical validation of the use of the Ratios as indicators of Group recognition will be given with these results. The procedure used in the validation provided additional interesting information, first in identifying a particular molecule format that elicited an unexpected response mode from students, and secondly in enabling a comparison to be made of the correct response rate for functional groups in different orientations. A general description of students' responses to Molecule Items will be given, and a more specific investigation comparing responses to items containing "back-to-front" and normally oriented functional groups will be reported. In this section, as in the previous one, the variation in mean scores across years will be considered in detail.

The final results section will be devoted to reporting and discussing the relation between the various Test scores and measures of performance in Chemistry.

Three matters will be considered before the presentation of the results. In Section 4.1, a statement of the Experimental procedure will be given, and the sample of pupils that participated in the Tests will be described. The second Section considers briefly the statistical analyses to be used in the Chapter.

4.1 Experimental Procedures

The procedures adopted for administering the Combined Tests were largely determined by the necessity for giving the Tests after the Organic section of the Chemistry course had been completed. As mentioned earlier, the Organic work is customarily the last part of the syllabus to be taught, and there was, therefore, a very short space of time available for carrying out the experiment in all schools before pupils' departure for public examinations. For this reason, Chemistry Departments were asked to administer the Combined Tests to their own pupils if this were possible.

In November 1973 the Chemistry Departments of twenty-six schools were asked if they would be willing to participate in the experiment.

At that time, the teachers' co-operation in administering the Tests was requested, and an estimate of the number of pupils they would be able to involve was also sought. Staff members were informed that the Test items would be provided in the form of sets of overhead transparencies and/or sets of large Test cards.

Nineteen schools indicated that they were willing and able to participate in the experiment, and in each case, the Staff members willingly offered to give the Tests themselves. Sets of Test materials and answer sheets were distributed to schools in February, 1974, to allow teachers the maximum choice in determining a convenient time for administering the Tests. A set of instructions detailing the procedure to be followed was also provided. In almost every case, it was possible to supplement this instruction with a personal visit to a Department, during which the Experimental procedure was discussed in detail.

A copy of the instruction sheet is attached in Appendix 4.1. Briefly, the administration of the Combined Tests followed the procedure described in the previous Chapter (p. 78). It seemed possible that the younger pupils could be affected by fatigue during the Molecule Test, in spite of the prescribed rest period preceding it. One would then expect their Molecule Scores to be depressed somewhat, and thus their Ratio Scores would be lower and could underestimate their ability to recognize groups as units.

In principle, it would have been possible to obtain an estimate of such an effect by using a "counter-balanced" Experimental design, in which each school divided their participating pupils randomly into two groups, which would be given the Tests in opposing orders. For pupils given the Molecule Test first, one would expect fatigue effects to result in an artificial lowering of the Visual Score, and consequently a spurious raising of the Ratios. Thus, the fatigue effect would act in opposite directions for Pattern first and Molecule first groups, and so a test for no significant difference between group results within each year would provide a sensitive test of the fatigue factor.

Apart from the fact that such a procedure would have placed a double burden on teachers, two theoretical difficulties were apparent. First, the division of pupils would have had to have been "random" with respect to Visual Ability and the ability to recognize groups as units - both unknown quantities. This difficulty could have been mitigated by making the split random with respect to age and performance in Chemistry - the best practical (and arguably reasonable on theoretical grounds) approximation to the required split. However, the second difficulty seemed crucial. If pupils were given the Molecule Test first, they could later recognize the Patterns as formula-analogues. This could enable them to "chunk" the patterns, and could therefore result in an increase in Visual Scores - and possibly a very large increase. It seemed that this effect could introduce a much greater error into the Experimental results than the fatigue factor, and therefore the Pattern first order was specified for all pupils. It was also expected that any serious effect due to fatigue would become apparent in the validation of the Ratios as indicators of group unit recognition (Section 4.41).

4.11 Description of the Sample

Sixteen schools were finally able to participate in the Experiment. A detailed description of the sample, contained in Appendix 4.2, indicates the types of schools involved, and the distribution of participating pupils within them.

A total of 1361 pupils were involved. The Test results from 69 pupils were not included in the analyses. (One group of 20 fifth year pupils were not shown one of the Patterns, and 26 fourth year pupils from one school appeared not to have finished the Pattern Test; the remainder were withdrawn because of spoiled scripts).

The distribution of pupils providing the 1292 sets of results that were used in the analyses was:

6th Year SYS	6th Year 'H' Grade	5th Year 'H' Grade	4th Year 'O' Grade
49	119	427	697

This gave a sample ratio of 1:11:14 for SYS:'H':'O' presentations, which compared favourably with the overall Scottish population ratio of 1:9:17 for the 1974 presentation cohort. From the point of view of statistical analysis, however, the number of SYS pupils was rather low. A larger sample of SYS pupils had been expected because of an imprecise question in the first communication with teachers, which asked for an estimate of the number of Sixth Year pupils who could participate. It was assumed that the majority of these would be SYS candidates. The figures above show that this assumption was erroneous.

4.2 Statistical Procedures

In this Chapter, as in Chapter 3, the confidence intervals for measurements will be reported whenever the test statistic applied enables the computation of such an interval. The way in which a confidence interval indicates the uncertainty in the value of a population parameter following measurement of a sample characteristic has been illustrated in Chapter 3 (p. 82). In this Chapter, as well as estimating population parameters, it will be necessary to compare pairs of estimated values.

Clearly, if the confidence intervals constructed about two sample values do not overlap, we would expect a statistical test to indicate that the corresponding population parameters differed significantly at a level comparable to that of the confidence interval (for example, at a level $\alpha \leqslant .05$, where 95% confidence intervals had been constructed). However, it should be noted that two parameters may be found to be significantly different, when the associated confidence intervals overlap to some extent. Therefore, when the difference between two estimated parameters is being considered, the value of p (the probability of obtaining a difference of the observed magnitude, where the mean population difference is of the postulated size) will be quoted in addition to the two confidence intervals. For consistency, the 95% confidence interval about each difference could be reported. However, the possible range of a difference (at the 95% level) is indicated almost exactly by the confidence intervals about the sample values; reporting "p" is, in effect, a more concise way of indicating whether or not the difference range includes the posited value.

This particular method of reporting statistical analyses is not widely used as yet, and therefore it would seem appropriate to give a brief apologia for its use in place of the more frequently employed Null Hypothesis Testing procedures.

Since the 1950's, Null Hypothesis Testing procedures have been the subject of severe criticism on several grounds, one of which - the effect of sample numbers on test results - is particularly relevant for this study. The way in which sample size influences the interpretation of results has been described differently according to the point of view of the experimenter. Grant,[63] arguing within the

Fisherian[64] school, has stated:

> "The tactics of accepting H_o as proof and rejecting H_o as disproof of a theory lead to the anomalous results that a small-scale, insensitive experiment will most often be interpreted as favouring a theory, whereas a large-scale, sensitive experiment will usually yield results opposed to the theory!"

On the other hand, Meehl[65] has stated:

> "In the physical sciences, the usual result of an improvement in experimental design, instrumentation, or numerical mass of data is to increase the difficulty of the 'observational hurdle' which the physical theory of interest must successfully surmount; whereas in Psychology and some of the allied behavioural sciences, the usual effect of such improvement in experimental precision is to provide an easier hurdle for the theory to surmount". (Emphasis added).

These two arguments appear to contradict each other, in that Grant has suggested that a large sample size will increase the likelihood of a theory's being rejected, while Meehl argues that a large sample size makes acceptance of a theory more likely. This contradiction is only apparent; Grant's 'theory of interest' is the Null Hypothesis (H_o), whereas Meehl's is the Alternative Hypothesis (H_1). The two quotations are simply different ways of saying that small sample size favours acceptance of the Null, while large sample size favours the acceptance of the Alternative Hypothesis.

Examination of any of the well known 'Null Hypothesis' formulae shows that this must be so. Consider, for example, the normal procedure for comparing the Means, M_a and M_b of two independent samples (each of size n, and having variances s_a^2 and s_b^2). Two hypotheses will be proposed:

$$H_o \; : \; M_a - M_b = 0$$

$$H_1 \; : \; M_a - M_b \neq 0$$

(It should be noted that while it is preferable to propose a specific value of $M_a - M_b$ for H_1, the inspecific formulation given above is generally used).

A t-statistic will be computed with 2n - 2 degrees of freedom, via the formula

$$\frac{|Ma - Mb|}{\sqrt{(S_a^2 + S_b^2)/n}} = t,$$

and the means would be considered "significantly different" (at $\alpha = .05$) - that is, H_1 would be accepted - if the computed value were greater than the 95th percentile value of t^{2n-2}.

Putting this another way, the means would be 'significantly different' at ($\alpha = .05$) if

$$|M_a - M_b| \geq \frac{t^x}{\sqrt{n}} \sqrt{s_a^2 + s_b^2}$$ (where t^x is the appropriate 95th percentile value).

The term that depends on n, $\frac{t^x}{\sqrt{n}}$, will decrease as n increases, and therefore the difference required for "significance" will also decrease with increasing n. It is generally true that, for all such tests where the sample size is large, a very small difference in sample values (or a very low value of, say, a correlation coefficient) will be found to be "significantly different from 0".

Such a "significant difference" may be of no practical consequence - for example, Cohen's tables[66] show that for a sample of 600, a correlation of r = 0.08 is "significantly different from 0" (at $\alpha = .05$). Therefore, if "significant difference" is the only criterion considered, a large sample size will bias an experiment in favour of rejection of the Null, and hence acceptance of the Alternative. Nunnally[67] and Binder[68] have stressed this point, and it is to this situation that Meehl referred in the quotation given above. On the other hand, where sample size is low, a gross difference must exist before a "significant difference" will be found. In this case, there is a numerical bias in favour of the acceptance of the Null, as Grant describes.

Of course, the latter problem can be overcome to some extent by quoting the Type II error entailed in accepting the Null hypothesis in favour of a specific alternative. It is worth noting, however that quotation of a Type II error is the exception rather than the rule;[69,70] and Type II errors are, of course, irrelevant where the numerical bias is in favour of rejecting the Null. In the present study, given the large sample sizes, it is the latter bias which would generally be of concern.

The particular argument outlined here against Null Hypothesis testing is that an experiment may be biased in favour of the "physical theory of interest" simply by choosing an appropriate sample size. It should also be noted that, where sample size is fixed, it is often possible to reformulate H_o and H_1 so that the numerical bias will lie in the "desired" direction. For instance, Grant[63] discusses the testing of a set of predicted values and observed values; in this case H_o could be "exact correspondence between theoretical and empirical points" (i.e. difference = 0). However, as accepting the Null is philosophically not de rigeur in the Fisherian school, Grant proposed a solution that involved reformulating H_o.

> "Basically the statistical argument in the proper test is reoriented so that rejection of H_O constitutes evidence favouring the theory. The new H_O is that the correlation between the predicted values ... and the obtained values ... is zero ...".

While this solution was proposed on ideological grounds, the same principle may be employed in a "numbers game". Rozenboom[71] describes a hypothetical, small sample experiment whose results could support a theory of no difference between means, or a theory of "the difference between means is 10", depending on the "orientation" of H_o (and a careful neglect of Type II errors).

In one way, it could be argued that the numerical bias described is not so much an inherent weakness of the Null Hypothesis testing procedures, as it is a property that exacerbates problems caused by the apparent willingness to let statistics make our decisions for us, and furthermore to allow these decisions to be made on an all-or-none basis.[71,67] To paraphrase De Rújula,[72] there seems to be a widespread tendency to use statistics as a drunkard uses lamposts (for support rather than illumination).

While confidence intervals are calculated using the same test statistics and distributions as null hypothesis tests, they are not susceptible to numerical bias, nor can they act as decision makers by default. Nunnally[67] put the matter succinctly:

> "The statistical hypothesis testing models differ in a subtle but important way from the confidence methods. The former make decisions for the experimenter on an all-or-none basis. The latter tell the experimenter how much faith he can place in his estimates, and they indicate how much the N needs to be increased to raise the precision of estimates by particular amounts".

Some examples will show how confidence intervals are not prone to numerical bias. A result of "r = .5, significant at the α = .05 level" for a sample of N = 30 would often be said to "support H_1". Reporting the confidence interval, {.16,.73} for the same results makes it clear that in fact the only information given by the experiment is that the correlation is probably positive, but of indeterminate magnitude.

Where a small N sample interval captures zero, the width of the interval will indicate very clearly the inadvisability of accepting H_o with any certainty. Where large N samples are concerned, a small difference of say .15 could be significantly different from zero. Reporting a confidence interval of {.1,.2} places the onus on the experimenter to delineate the extent to which this range of values could support, or fail to support, the physical theory of interest. The latter example also shows that increasing the sample size gives increased precision, by narrowing the confidence interval (or decreasing the uncertainty in the estimated parameter), but does not provide an easier hurdle for a theory to surmount (Meehl, op. cit.).

In summary then, as Rosenboom has said:(71)

> "The confidence interval report is not biased toward some favoured hypothesis, as is the null-hypothesis significance test but it makes an impartial simultaneous evaluation of all the alternatives under consideration".

The confidence interval is so clearly a description of the results that the experimenter must set forth his grounds for claiming the results as evidence in support of his hypothesis; he cannot claim "significantly different, Q.E.D.". Finally, in addition to these theoretical advantages, confidence intervals provide a complete, but concise description of results for each reader, who can then make an independent critical assessment of the experimental conclusions.

4.3 The Pattern Test Results

4.31 The Pattern Test Responses

To obtain the most complete description of responses to the Pattern Test Items the results obtained from pupils in all years were considered together initially. (This procedure was valid whether or not pupils in different years differed in Visual Ability, as such a

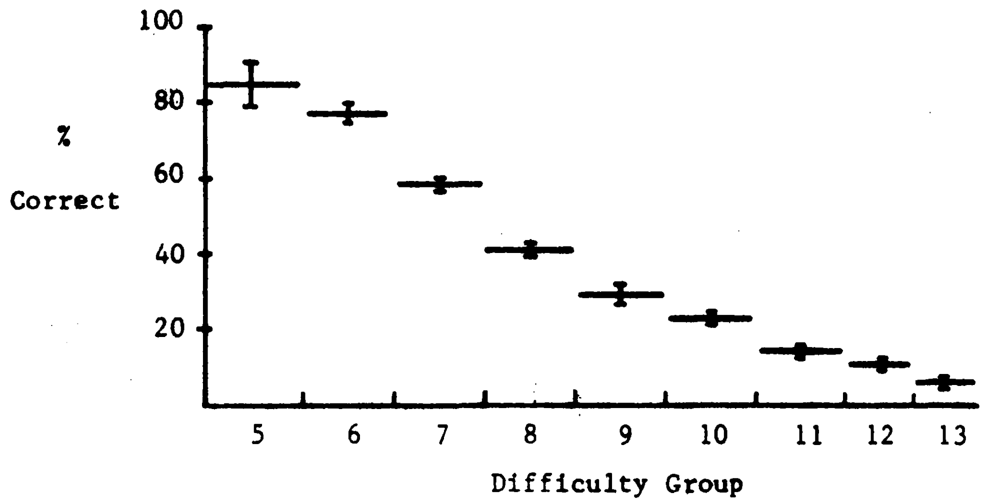


Fig. 4.1 Mean Correct Responses (as percentages) for each Difficulty Group (N = 1292).
The Error Bar indicates the S.E.

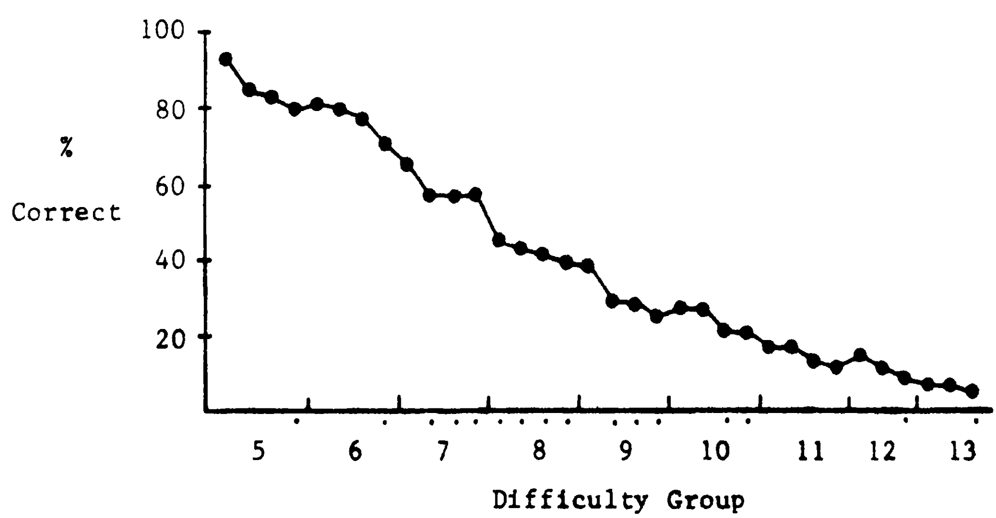


Fig. 4.2 Percentage Responses Correct for Each Item (N = 1292)
. Indicates an (integer + ½) Item

difference would affect only the rate of decline of correct responses per Item across Difficulty Groups).

The raw data have been reported as percentages of responses that were correct so that comparisons between the combined results and the results of individual years are not complicated by the different sample sizes. The mean percentages of responses correct for each Difficulty Group are shown in Figure 4.1, where error bars indicate the Standard Errors of the means. A fairly sharp drop between successive means is apparent from Groups 6 to 9, with a decreased rate of decline for the higher Groups. However, the fact that each Group mean is lower than the previous mean demonstrates that the required monotonic relationship between response rate and difficulty obtained.

The percentages of responses correct for each item in the Pattern Test are shown in Figure 4.2. In considering these results, some criterion had to be adopted for distinguishing spurious or random fluctuations in correct response rates from possibly genuine differences.

The 95% confidence interval about the difference between two percentages P and P' is given by $(P - P') \pm W$, where W (the width of the confidence interval) is a function of P and P' as well as the sample size(s). Therefore, the width of a confidence interval is not constant for a particular sample. However, it is possible to calculate the maximum width a confidence interval may have for any given sample size(s) for a fixed value of α; clearly, if $|P - P'| > W_{max}$ the confidence interval will not capture 0, whatever the values of P and P'. The expression for W_{max} is derived in Appendix 4.3; for a sample size $N = 1292$, $W_{max} \simeq 4\%$. In the light of the discussion in the previous section, no claim is made that differences of 4% are necessarily of practical consequence. Nevertheless, the maximum width value provides a useful order of magnitude figure, below which differences may reasonably be considered to be random fluctuations, and above which differences would merit further consideration.

Two comparisons of responses to categories of Items were of particular interest. The first of these was the response rates for integer and (integer + ½) Items within a Difficulty Group. Inspection of Fig. 4.2 shows that Difficulty Groups 5,6,7,9,10,12 and 13 contained both integer and (integer + ½) Items, and that the percentage of responses correct for (integer + ½) Items was slightly lower than for corresponding integer Items. For the extreme Groups, the difference was about 2-3%, while for the intermediate groups the difference was of the order of 6-9%. The same trend was apparent in the separate results for Sixth, Fifth and Fourth year pupils (shown in Appendix 4.4). Overall, the difference between the two sets of Items was only just above the level of random fluctuations, suggesting that the combination of integer and (integer + ½) Difficulty Items within a Difficulty Group was a valid procedure.

The third Item in Difficulty Group 9, and the last two Items in Difficulty Group 11 had been classed as having "left hand confusion" (Section 3.41). The difference between the "left hand" Items in Group 11 and the other two Difficulty 11 Items was of the order of random fluctuations, and there was no apparent difference between the response to the Difficulty 9½ "left hand" Item, and the other 9½ items. This suggested that the addition of 1 unit to an Item's Difficulty Number adequately accounted for the "left hand" characteristic. This point will be considered further in Section 4.61.

The only Item that appeared to draw an anomalous response was the first item in Difficulty Group 5. Coincidentally, this was the first Test Item shown to pupils, and this may well account for the increased percentage of correct responses. It will be recalled that an additional pattern that did not have to be recorded was shown to pupils before beginning the Test proper; if this Test were given again, pupils could be asked to record the practice pattern, and this would determine whether the anomalous response was due to the Item, or merely to its particular position.

An inspection of the responses recorded by pupils showed the characteristics already reported for the trial tests, namely a universal tendency to record the long central chain first, and then side chains starting from the left hand end of the pattern. This suggested that pupils "read" the patterns almost as though they were two-dimensional words. Characteristically, incorrect responses were incomplete, but sensible representations of the original Items.

4.32 <u>Visual Scores</u>

The mean Visual Score and 95% confidence interval for each group of students is given in Table 4.1, together with the probabilities, p, that differences of the observed magnitudes would have been obtained from pairs of samples drawn from the same population.

The reported figures did not suggest a clear trend of increasing Visual ability with age. While there was an increase of approximately 0.5 between the Fourth and Fifth year means, the difference between Fifth and Sixth Year Studies pupils would not normally be considered significant. (It can be seen that the small SYS sample size was associated with a large confidence interval). Furthermore, the Sixth year 'H' interval coincided with that of the Fourth years. In fact, when the results of all Sixth year pupils (who formed a group of comparable age) were combined, the three Age Group means fell within the approximate range 7.5-8.0, with the Sixth year mean at the mid-point of the range. These three results not only failed to show a unidirectional change in Visual Score with age, but also showed that the probable variation in Visual Score over the age range tested was very small in practical terms.

TABLE 4.1

MEAN VISUAL SCORES FOR AGE AND EXAMINATION GROUPS

Number	Group	Mean	95% Confidence Interval	Probability
49	Sixth Year SYS	8.47	{7.9, 9.0}	p < .001 (49 vs 119)
119	Sixth Year 'H'	7.44	{7.2, 7.7}	
168	Sixth Year Total	7.74	{7.5, 8.0}	p < .1 (168 vs 427)
427	Fifth Year 'H'	8.0	{7.8, 8.2}	p < .001 (427 vs 697)
697	Fourth Year	7.48	{7.4, 7.6}	

The difference between Fourth and Fifth year means would suggest that on average Fifth year pupils were able to reproduce correctly Patterns containing one more dot than those within Fourth years' competence. The difference between the most extreme groups - the Fourth year pupils and the Sixth year SYS pupils - corresponded to a pattern difference of only one or two dots. Another practical measure of the Visual Abilities represented by the mean Visual Scores may be obtained by considering the complexity of formula that could be reproduced correctly, given the mean Visual Score of each group of pupils. Providing that he recognised the Class III groups as units, a pupil having Visual Score 7.5 or 8.0 could be expected to reproduce formulae equivalent in complexity to a simple ester (such as methyl propanoate), while, under the same conditions, a Visual Score of 8.5 would indicate an ability to reproduce say, ethyl propanoate correctly.

In practical terms, therefore, the results suggested that there was no material difference in Mean Visual Ability of the four groups tested, and indicated that there was possibly not even a directional increase in mean Visual Ability with age.

Table 4.2 shows the mean and standard deviation of Visual Score for each group, and also the mean and standard deviation obtained when all results were combined. It is very instructive to compare these figures with the Short Term Memory Capacity of 7 ± 2 chunks of information (Section 3.3). Clearly, the overall result of 7.7 ± 1.6 and the Memory capacity of 7 ± 2 are numerically equivalent; however, it would be unwise to infer from this that a pattern unit awarded a Difficulty score of 1 necessarily represented 1 chunk of information. It will be recalled that a repetition of a dot or shape scored only ½; this could not be equated to a "half chunk", as chunks are, by definition, indivisible. Nevertheless the Difficulty Score system and a chunking system are not incompatible. For example, a unit of 5 dots on a line would be awarded a Difficulty Score of 4, and it certainly could be stored as 4 chunks; <u>a dot</u>, <u>two more</u>, <u>two more</u>, <u>on a line</u>. The principle underlying the construction of the Difficulty Number system was that the Difficulty of a Pattern depended on the number of things that had to be noted about it, and a repetition was scored as being less difficult because it was assumed that less information would need to be recorded about it (i.e. it contained less information). Given this underlying principle, the numerical coincidence of both the overall mean and the standard deviation suggested two interesting implications about Visual Ability.

TABLE 4.2

MEAN VISUAL SCORES

Group	Mean	S.D.
Sixth Year SYS	8.5	1.8
Sixth Year 'H'	7.4	1.5
Fifth Year 'H'	8.0	1.7
Fourth Year 'O'	7.5	1.5
Combined	7.7	1.6

First, it would seem that the mechanisms required (perceptual, coding, processing for retrieval, etc.) to store and retrieve this type of pattern did not present a greater hurdle to these pupils than those required for other types of information - number or letter strings, for example. If the processing of this type of data had, in some way, been particularly difficult, we would have expected a much lower mean Visual Score. Secondly, it would seem that the ability to reproduce Patterns correctly was limited by the capacity of Short Term Memory rather than by Visual Ability directly.

It is worth noting that, apart from any inferences made about Visual Ability and Memory Capacity, the overall mean and standard deviation of 7.7 ± 1.6 indicated the "Pattern Capacity" range of Short Term Memory; an application of this range will be discussed in Chapter 5.

4.4 The Molecule Test and the Combined Test Results

4.41 The Validation of the Ratios as Determiners of Group Recognition

Before considering the Molecule Test and Combined Test results, it is necessary to establish that the Class to which a pupil was assigned on the basis of his Class I-III ratios validly indicated the set of Groups he recognised as units.

In analyzing the results of the Combined Tests trial the set of correct responses expected of a student on the basis of his Visual Score and assigned Class was compared with his actual set of correct responses. While this procedure, if used in analysing the present results, would have shown up gross discrepancies, it was difficult to see how it could have been used to provide a critical test of the validity of the Ratio assignment system, because one could make only an arbitrary specification of the degree of coincidence between an expected set and an observed set that would be considered "acceptable", and the proportion of pupils having "coincident" sets that would be accepted as being indicative of validity.

The large sample size, however, allowed the use of another procedure, that was susceptible to a stringent statistical analysis. Instead of considering the responses that each student should have recorded correctly, we determined which students should have recorded each Molecule Item correctly. Consider a Molecule Item having Difficulty 4,6,7,9, when Class I, II, III and IV groups respectively were scored as single units. This Item would be reproduced correctly by the following groups of pupils:

Class I pupils having Visual Score $\geqslant 4$

Class II pupils having Visual Score $\geqslant 6$

Class III pupils having Visual Score $\geqslant 7$

Class IV pupils having Visual Score $\geqslant 9$

providing that each pupil recognised as units just those groups corresponding to his assigned Class.

In principle, then, the categories of pupils expected to reproduce each Item correctly were determined in this way, and a two-way classification of pupils by Visual Score and assigned Class enabled the number of correct responses expected for each Item in the Molecule Test to be computed. In practice, a slightly more detailed procedure was used to determine the expected numbers, to take account of the fact that the Visual Score system did not require all Items within a "correct" Group to be reproduced correctly, and also to allow for non-integer Visual Scores. This detailed procedure is described in Appendix 4.5.

Predicting the number of correct responses expected for an Item simultaneously predicted the number of incorrect responses to be expected; that is to say, the predictions made were of the proportion of responses expected to be correct, rather than the frequency of correct responses. In the same way, the observed totals indicated the observed proportion of responses that were correct. The expected proportions, P_e, and the observed proportions, P_o, were thus compared.

Formally, we tested H_o: $P_e = P_o$ against H_1: $P_e \neq P_o$ for each item via the z statistic[(73)]

$$\frac{P_e - P_o}{\sqrt{(p(1 - p)2/N)}} = z.$$

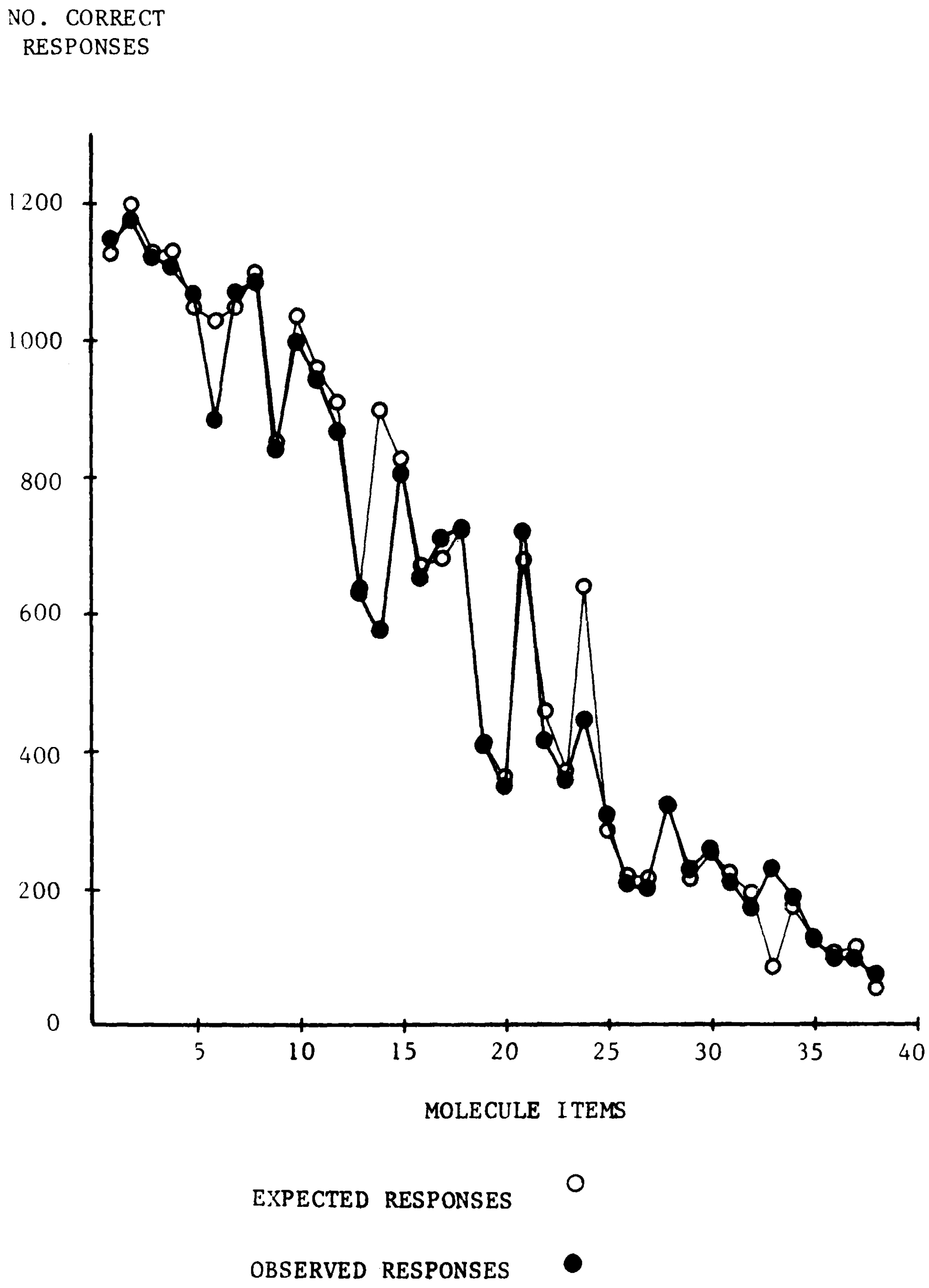
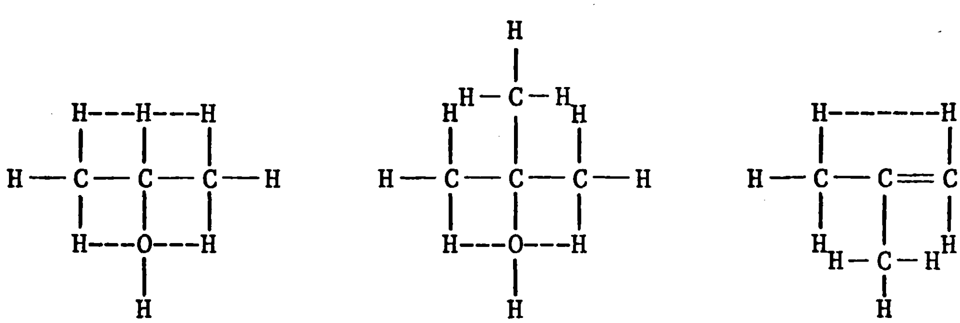


Figure 4.3 **Expected and observed** numbers of correct **responses** for each Molecule Item

The power for each such test was greater than .99, for a small difference between P_e and P_o. By squaring the z statistic,[74] a χ_1^2 statistic was obtained. Each Item in the Molecule Test then formed a replication of the test of H_o against H_1, and the individual χ_1^2 values were summed to give χ_n^2, where n was the number of Test Items. This chi-squared test formed an extremly stringent test of H_o, and hence of the validity of the assumption that a student did recognize as units those groups corresponding to his assigned Class.

The expected and observed numbers of correct responses for each Item, presented in Figure 4.3, showed very close agreement for 34 of the 38 cases. It can be seen that the expected number greatly exceeded the observed number of correct responses for Items 6, 14 and 24. These Items had a common pattern characteristic, and were termed "box molecules", because of the very frequent and idiosyncratic incorrect response pattern given for them. These three items are reproduced in Figure 4.4, and the additional lines that were characteristically found in incorrect responses are shown as dotted lines.



Item 6 Item 14 Item 24

Figure 4.4 The three "Box Molecules"

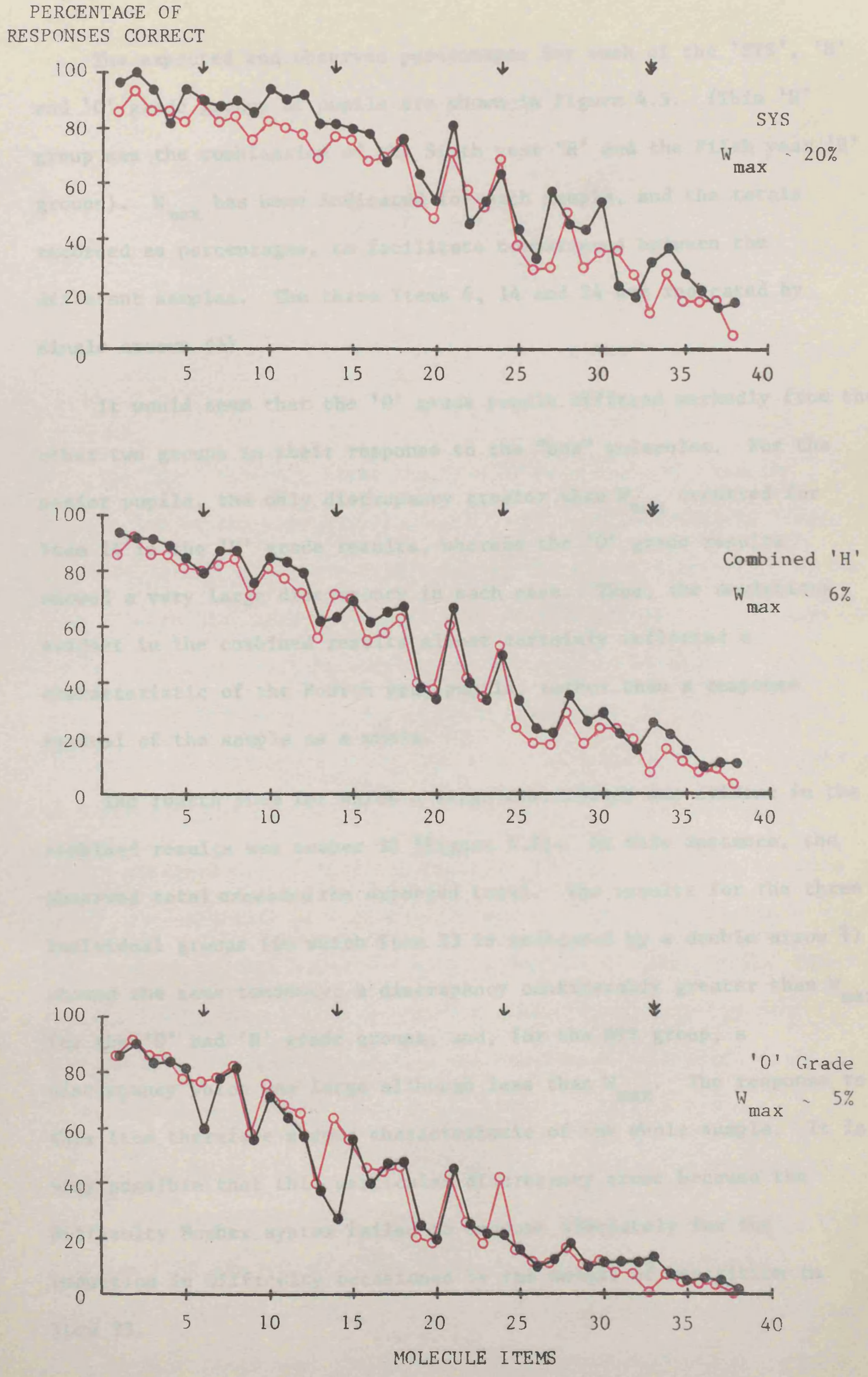


Figure 4.5 The Results of Figure 4.3 shown as percentages for the three examination groups

The expected and observed percentages for each of the 'SYS', 'H' and 'O' grade groups of pupils are shown in Figure 4.5. (This 'H' group was the combination of the Sixth year 'H' and the Fifth year 'H' groups). W_{max} has been indicated for each sample, and the totals recorded as percentages, to facilitate comparisons between the different samples. The three Items 6, 14 and 24 are indicated by single arrows (↓).

It would seem that the 'O' grade pupils differed markedly from the other two groups in their response to the "box" molecules. For the senior pupils, the only discrepancy greater than W_{max} occurred for Item 14 in the 'H' grade results, whereas the 'O' grade results showed a very large discrepancy in each case. Thus, the deviations evident in the combined results almost certainly reflected a characteristic of the Fourth year pupils, rather than a response typical of the sample as a whole.

The fourth Item for which a large discrepancy was evident in the combined results was number 33 (Figure 4.6). In this instance, the observed total exceeded the expected total. The results for the three individual groups (in which Item 33 is indicated by a double arrow ↡) showed the same tendency; a discrepancy considerably greater than W_{max} for the 'O' and 'H' grade groups, and, for the SYS group, a discrepancy which was large although less than W_{max}. The response to this Item therefore seemed characteristic of the whole sample. It is very possible that this particular discrepancy arose because the Difficulty Number system failed to account adequately for the reduction in Difficulty occasioned by the amount of repetition in Item 33.

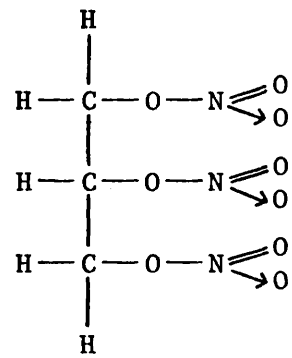


Figure 4.6

The "box" Molecule Items and Item 33 were excluded from the computation of the overall value of χ^2. For the remaining 34 Items, χ^2_{34} = 34.7, the 55th percentile value of the χ^2_{34} distribution. As the type II error for each individual test of proportions was less than .01 for a small difference in proportion, the overall value of χ^2 gave very good grounds for accepting H_o: $P_e = P_o$, and hence for accepting the assumption that students recognized as units those groups corresponding to their assigned Class.

A possible explanation for the fourth years' anomalous response to "box" molecules will be suggested in the next section, and the implications of this behaviour will be considered later in the Chapter. At this point, however, it is worth considering whether the four discrepant Items could invalidate the conclusion just reached, on technical grounds.

First, it should be noted that as far as the value of his Class I-III ratios was concerned, the index of a student's performance in the Molecule Test was his Molecule Score, not the number of correct responses made. Therefore, the ratio values would have been affected only to the extent that the Molecule Scores were affected by these anomalous responses. The Scoring system (Appendix 3.2) made an allowance for a "lucky chance" correct response; because Item 33 occurred in Difficulty Group 14, the second highest Difficulty Group, this property would have prevented some pupils gaining credit for an anomalous correct response for this Item. The Scoring system also defined Groups to be "correct" where no more than one mistake was made per Group in consecutive Groups; therefore, if pupils had the first 13 Groups "correct", and the other two Items in Group 14 correct, (for Class I-III pupils the latter would be an expected concomitant of the former) they would obtain no additional credit for an unexpected correct response for Item 33. The latter property of the Scoring system would also mean that pupils would not necessarily be penalised for an unexpected incorrect response for any or all of the "box" molecule Items, which occurred in different Difficulty Groups. (In particular, the value of the Class I ratio - the most important ratio - would have been expected to be susceptible to unexpected incorrect responses only for Item 14, and then only for students having Visual Score 6, given the particular Items in the Test).

Secondly, the classification of a pupil would have been affected only if any affect on his Molecule Score was sufficient to change the value of the ratios to the extent that his "correct" ratio was no longer the ratio nearest 1. The important question, therefore, was the robustness of the procedure for assigning pupils to Classes. The question may best be resolved by consideration of the results themselves; the accuracy of the predictions for the non-discrepant Items surely indicates that the Classification procedure was sufficiently sensitive to distinguish between the different recognition behaviours, but was, at the same time, sufficiently robust to withstand any affect due to the two types of anomalous behaviour observed. The results also indicated that any affect due to fatigue (Section 4.1) must have been within the robustness level of the procedure.

4.42 The Ratio Scores

Table 4.3 shows the percentages of pupils assigned to each Class - and therefore the percentages who recognised as units the corresponding groups. The most significant result here was that in the sample as a whole, only 4.6% of the pupils recognised as units the Class I groups - it will be recalled that such recognition behaviour was the proposed indicator of the first required level of conceptual understanding. Clearly, the vast majority recognised only the Class III groups as units while those recognizing no groups as units formed a substantial proportion of the sample.

The distributions of Sixth year 'H' and Fifth year 'H' pupils were, as might have been expected, fairly similar. For these two samples, W_{max} was 10%, indicating that while the proportions in Class III were possibly significantly different, there was no significant difference in any other category. Furthermore, the differences did not suggest that either 'H' grade sample tended to be uniformly better than the other. A combined distribution for the two 'H' grade samples was therefore computed, and has been included in Table 4.3.

This combined 'H' group differed from the SYS sample in having a non-negligible percentage of pupils in Class IV, but, more importantly, the percentage of pupils in Class I was the same for both groups.

The distribution of the Fourth year pupils was clearly different from those of the more senior pupils, particularly at the extremes of the distributions. An almost negligible proportion of Fourth year pupils was assigned to Class I, while a very substantial proportion

TABLE 4.3

PERCENTAGES OF PUPILS IN EACH CLASS*

Group	Class I %	Class II %	Class III %	Class IV %
Sixth Year SYS	8.2	20.4	71.4	-
Sixth Year 'H'	9.2	9.2	78.2	3.4
Fifth Year 'H'	8.7	16.2	67.9	7.3
Combined 'H'	8.8	14.7	70.1	6.4
Fourth Year 'O'	1.1	3.6	80.2	15.1
Combined	4.6	8.9	75.6	10.8

* The groups associated with each Class were:

Class I $-C(=O)-O-H$, $-O-H$, $>C=O$, $-CH_3$, $>CH_2$

Class II $=O$, $-O-H$, $-CH_3$, $>CH_2$

Class III $-O-H$, $-CH_3$, $>CH_2$

Class IV No groups recognized as units

fell within Class IV. The size of this latter group could perhaps account for the fact that only Fourth years' observed correct totals for the "box" Molecule Items deviated substantially from the expected figures. It would seem possible that a pupil who recognized $>CH_2$ and $-CH_3$ groups as units would be less likely to be distracted by the implicit "box" pattern of these Items; if this were the case, the Fourth year results would be expected to show a large anomalous effect as the proportion of pupils in Class IV was so much greater in this sample than in the other year groups.

The difference in performance across years was shown more concisely by the mean Class I ratio Scores for each year group (Table 4.4). The intervals for the 'H' grade samples showed almost complete overlap, and therefore an 'H' grade mean and interval was computed and used in comparing recognition behaviour across years. The Fourth year results were clearly much lower than the 'H' grade results, which were, in turn, significantly lower than the SYS results. It should be noted, however, that the separation of the SYS and 'H' grade intervals was not large (possibly due, in part, to the comparatively large uncertainty in the SYS result).

The combined results thus showed a trend towards better performance going from the 'O' grade to the SYS samples, with the greatest difference being found between 'O' grade and 'H' grade results. The most significant aspect of the results was, however, the very low proportion of pupils demonstrating recognition of the Class I groups as units, and, concomitantly, the very low mean Class I ratio Scores.

TABLE 4.4

MEAN CLASS I RATIO SCORES

Number	Group	Mean	95% Confidence Interval	Probability
49	Sixth Year SYS	.62	{.55, .69}	p < .01 (49 vs 119)
119	Sixth Year 'H'	.47	{.42, .53}	p < .5 (119 vs 427)
427	Fifth Year 'H'	.50	{.47, .53}	
546	Combined 'H'	.49	{.46, .52}	p < .001 (546 vs 697)
697	Fourth Year 'O'	.27	{.25, .29}	

4.43 An Examination of Responses

Before considering the results of the examination of responses, it is perhaps worth recalling the three types of information that were of interest in this aspect of the investigation. In the first place, we looked for indications of confusion in the perception or recording of Items. Secondly, it was necessary to look for any evidence of perception of Class I functional groups first, in the form of correctly recorded functional groups in otherwise completely incorrect responses, the criterion adopted for the second level of conceptual understanding of functional groups (Section 2.52). Lastly, because of the specific suggestion that pupils' difficulties were a result of "back-to-front" formulae, a direct comparison of the reproductions of functional groups in different orientations was made.

The unusual response format for "box" Items has already been mentioned, but apart from this, no indication of confusion was apparent in incorrect responses. Again, the incorrect responses appeared to be incomplete or inaccurate, but perfectly sensible, representations of the original Items. As in the case of the Pattern Test, the responses suggested that pupils had "read" the formulae from left to right, starting with the central chain, and had recorded them in the same order.

There were no instances noted of pupils recording a functional group correctly, and having the rest of the Molecule Item completely wrong. This was perhaps hardly surprising, in view of the very small number of pupils who were assigned to Class I. This recognition behaviour was proposed as the observable indicator of the conceptualisation of the functional groups as chemical entities, and one would certainly have expected that pupils with this conceptualisation would have been more likely to recognize functional groups first, than pupils who had not achieved that level of understanding. Therefore, it had seemed that the incorrect responses of Class I pupils would be the most likely source of evidence of recognition of functional groups first, if such evidence were to be found. There were, however, only a relatively small number of incorrect responses from Class I pupils - a Class I pupil, having a Visual Score close to the Mean of 7.7 would have been expected to give very few incorrect responses in the Molecule Test. Therefore, a large number of Class I pupils would have been required to obtain a viable subset of Class I pupils with low Visual Scores whose greater number of incorrect responses could have provided a better test-bed for the perception of functional groups first.

The $-\overset{O}{\overset{/\!/}{C}}-O-H$ group was the most fruitful functional group for study in this context, as it could be considered as part of the central chain, plus a side branch ($^{O}_{/\!/}$). It was the case that in the relevant incorrect responses of Class I pupils, this functional group was generally reproduced correctly. The behaviour suggested by these responses was that these Class I pupils also began by noting the long central chain first, but that for them this included the entire functional group, rather than just the elements -C-O-H - which was no more than the expected consequence of their recognition of the functional group as a unit.

The behaviour of non-Class I pupils was much more clear-cut. They clearly noted the long central chain first, and very often reproduced $-C^{\overset{O}{/\!/}}$ -O-H incorrectly. The three most common misrepresentations of this unit were: -C-O-H, $-C-\overset{\overset{O}{\|}}{O}-H$, and $-\overset{\overset{H}{|}}{\underset{\underset{H}{|}}{C}}-O-H$. That is, the -C-O-H was treated as part of the central chain, and the $\overset{O}{\|}$ as a side chain. Where a functional group was recorded correctly in an incorrect response, it was almost always the -O-H group - which was reproduced as part of the central chain, or as part of a side chain.

Overall, then, the evidence suggested strongly that pupils noted the central chain first, rather than any functional group, and that the Class I pupils very possibly differed in what they perceived as belonging to the central chain.

"Back-to-front" and "normally oriented" representations are just two of several possible orientations of functional groups. Although the functional groups -OH and $-C^{\overset{O}{/\!/}}$ -O-H appeared in several different orientations in various Test Items, it was not appropriate to draw inferences from direct examination of responses in most cases because of the differing Difficulties of the Items. Three Items were chosen for direct study; Items 15 (containing H-O-) and 16 (containing -O-H) which were of the same Class IV Difficulty, and very similar Class III

Difficulty, and Item 20 which contained a 'left-handed' and a 'right-handed' $-\overset{O}{\overset{\|}{C}}-O-H$ group.

An analysis was made of the responses given to these Items by every sixth pupil. The results obtained from the 202 selected answer sheets are given in Table 4.5.

TABLE 4.5

RESPONSES TO FUNCTIONAL GROUPS IN LEFT AND RIGHT HANDED ORIENTATIONS

	Only Right Hand Group Correct	Only Left Hand Group Correct	Both Correct	Two Identical Wrong Responses	Two Different Wrong Responses
OH Group	12	22	163	5	-
COOH Group	5	18	51	104	24

In the majority of cases, pupils gave the same response for the left-handed and the right-handed versions of a group. The numbers giving a correct response for only one of the representations of a group did not differ significantly (for N = 202, $W_{max} \sim 10\%$, i.e. 20 responses). If a trend was suggested by the single correct group responses, it was that the group occurring at the left hand end of the Item (the back-to-front group) was the more likely to be reproduced correctly. This was not unexpected, given the strong tendency to read an Item from left to right, and reproduce it in the same order. There was certainly nothing to suggest that the back-to-front or left-handed representations were intrinsically more difficult than the normally oriented versions of the groups.

The system used for assigning Difficulty Numbers to Items assumed that if a group was recognized as a unit it would be so recognized whatever the orientation of the group. Therefore, we would expect the accuracy of the prediction of numbers of correct responses for Items containing a particular group to be independent of the orientation of that group, if in fact pupils perceived all orientations as being equivalent in difficulty. As the prediction procedure took account of of the Difficulty of each Item, a comparison of expected and observed totals for sets of Items containing a specific functional group enabled a more extensive investigation of this facet of pupils' behaviour than a direct examination of recorded responses.

Items 15, 16, 25 and 26 contained the hydoxyl group in the orientations H-O-, -O-H, $\overset{\text{H}}{\underset{|}{\overset{|}{\text{O}}}}$, $\overset{\text{O}}{\overset{|}{\text{H}}}$, respectively. Items 5, 9, 11, contained the carbonyl group as >C=O, O=C<, and $-\overset{\text{O}}{\overset{\|}{\text{C}}}-$. Items 20, 32 and 38 each contained one left-handed and one right-handed $-\overset{\text{O}}{\overset{/\!/}{\text{C}}}-\text{O}-\text{H}$ group. Examination of Figure 4.3 shows that the agreement between observed and predicted totals was uniformly good within each of these sets.

Thus, neither pupils' actual responses, nor the comparison of expected and observed totals of correct responses showed any indication that pupils responded differently to a functional group in different orientations.

4.5 Correlations Between Test Variables and Performance in Chemistry

The comparisons of Visual Score means and Class I Ratio means across years formed one part of the investigation of the relationships between Visual Ability and performance in Chemistry, and between Conceptual Understanding of functional groups (recognition of groups as units) and performance in Chemistry. The second part proposed (Sections 2.51 and 2.52) was an examination of the correlations between the two Test Variables - Visual Score and Class I Ratio - and a measure of performance in Chemistry; the reporting and discussion of these correlation coefficients forms the main part of this Section.

First, however, the other correlation coefficients computed - those between Visual Score and Class I Ratio - will be considered. While the Pre-Test results had shown no evidence of a substantial linear relationship between these two variables, the validity of inferring that groups were not conceptualised as units from their not being recognized as units (Section 3.3) was dependent on the lack of any such relationship, and therefore it was felt important to examine this relationship again, given the more precise data available from this much larger sample.

The Pearson Product-Moment correlations between Visual Score and Class I ratio (and the associated 95% confidence intervals) for the four pupil groups are shown in Table 4.6.

TABLE 4.6

CORRELATION COEFFICIENTS, VISUAL SCORE - CLASS I RATIO

Group	r	95% Confidence Interval	β
Sixth Year SYS	-.05	{-.32, +.22}	.99
Sixth Year 'H'	-.06	{-.24, +.12}	> .995
Fifth Year 'H'	-.22	{-.32, -.12}	> .995
Fourth Year 'O'	-.18	{-.26, -.10}	> .995

The values of β quoted are the probabilities (for the different sample sizes) of obtaining a non-zero coefficient, if the population r were equal to 0.5. Clearly, the intervals for both Sixth year samples capture 0. The Fourth and Fifth year results suggested a negative and weak correlation between the two variables. There was thus no evidence to suggest a significant positive relation between Visual Score and Class I Ratio, replicating the finding of the Pre-test.

The other correlations to be discussed involved the relationship between a Test variable and a measure of performance in Chemistry. The most appropriate measure of performance would have been obtained on a specially devised Organic Chemistry test, administered to all participating pupils at approximately the same time as the Combined Tests. As insufficient time was available in schools for the administration of such a test, a measure of pupils' general performance in Chemistry was obtained from teachers, in the form of the rank order list describing their pupils' relative performance in their "Preliminary Examinations". In most cases, a separate list was provided for each form, although in a few instances a common rank order list was given for a complete presentation cohort. At best, then, the rank of pupils allowed a comparison of Test and Chemistry performances for pupils within individual forms or years in each School. However, a related categorization was available which permitted pupils from different Schools to be grouped together; this was a pupil's position "above" or "below" the Red Line. (In Scotland, teachers are required by the Examination Board to furnish an Order of Merit list for all pupils presented for the 'O' and 'H' Grade public Examinations, and to indicate, by a red line, the rank above which pupils are confidently expected to pass the appropriate Examination).

The fact that the rank order and Red Line categorisations of pupils related to their performance in Chemistry overall, rather than their performance in Organic Chemistry specifically, was not felt to be a serious disadvantage. There was no evidence that pupils of the ages considered showed marked differences in performance in the different branches of Chemistry. Scottish Sixth Year pupils may sit a public examination, known as a Bursary Examination to compete for a University Scholarship. An examination of 100 scripts from the 1973 Chemistry Bursary Examination showed an unbiassed correlation of 0.8 (95% confidence interval {0.6, 1.0}) between Organic marks and Total marks.

With one exception, Schools provided rank order lists for their participating pupils, and 11 Schools were also able to provide their Red Line categorisations. The Biserial Coefficients computed for "Visual Score" - "Above or Below Red Line" and "Class I Ratio" - "Above or Below Red Line" are shown in Table 4.7.

TABLE 4.7

BISERIAL COEFFICIENTS

	Visual Score - Above or Below Red Line	Class I Ratio - Above or Below Red Line
Sixth year 'H'	0.16 ($p < .3$)	0.10 ($p < .6$)
Fifth year 'H'	0.17 ($p < .02$)	0.26 ($p < .001$)
Fourth year 'O'	0.19 ($p < .001$)	0.22 ($p < .001$)

A coefficient could not be computed for SYS pupils, as teachers are not required to provide an Order of Merit list, with its associated Red line, for their Public Examination. As no technique is available for computing confidence intervals about Biserial Coefficients, only the values of the probabilities, 'p', have been quoted. Inspection of the quoted values shows that while four of the six coefficients differed from zero (at $\alpha = .05$) all the values were very low, and that there was no practical difference apparent between the two sets of coefficients. In spite of this similarity, consideration of the Visual Score results (Section 4.32) and the Ratio Score results (Section 4.42) suggested that a very different interpretation should be placed upon the two sets of coefficients.

It has already been suggested that the ability to reproduce patterns correctly was limited by Short Term Memory Capacity, rather than Visual ability per se; there was certainly a good spread of Visual Scores (Table 4.A5.1, Appendix 4.5). Under these circumstances, if a significant relationship between Visual Ability and Chemistry performance had existed, one would have expected that even the rather gross comparison afforded by the Biserial coefficient correlation method would have returned non negligible values of the coefficients. It therefore seemed valid to infer from the low coefficients obtained that no substantial relationship obtained between Visual Ability and performance in Chemistry.

In the case of the Class I Ratio, however, we have seen that only some 5% of the total population were assigned to Class I. The mean Class I Ratios showed that the majority of pupils in the 'O' and 'H' Grade groups fell far short of the criterial performance - so far short that we could justly classify them as being equally bad at recognizing Class I groups as units, in spite of the spread of Class I ratio Scores.

One would expect an association between conceptual understanding and performance in Chemistry; thus a substantial correlation between Class I ratio and Chemistry performance would have tended to validate the choice of the ability to recognize Class I groups as units as a criterion of the specified level of conceptual understanding of functional groups, in that it would have suggested a one-to-one correspondence between recognition of Class I groups as units and their conceptualisation as such. It has already been verified that the Class to which a pupil was assigned validly indicated the groups recognized as units; the Biserial correlation procedure had been proposed in order to obtain an independent test of the assumption that the Class I Ratio (or the Classification procedure) also validly indicated the attainment or otherwise of the specified level of conceptual understanding of functional groups.

If this assumption were true, because the majority of pupils had to be classified as equally bad at recognizing Class I groups as units, it would follow that they would have to be classified as equally lacking in attainment of the required level of conceptual understanding of functional groups. Under these circumstances, one would not expect to observe any marked correlation between Class I Ratio and performance in Chemistry.

On the other hand, if pupils who did not recognize Class I groups as units had nevertheless attained the specified level of conceptual understanding, no correlation between Class I ratio and performance in Chemistry would be expected, because there would not be a one-to-one correspondence between recognition as units and level of conceptual understanding.

Thus, the obtained distributions of pupils across Classes I-IV indicated first, that the Correlation procedure could not distinguish between these two cases as had been intended, and secondly, that the very low coefficients obtained were all that could be expected. The results have been quoted only for completeness, as the procedure had been specified in the Experimental Design.

Because the results indicated that the Correlation procedure was not appropriate under the extant conditions, two further investigations, not proposed in the Experimental Design, were made. The first consisted of a comparison of the percentages of "Class I" pupils and "not Class I" pupils above the Red Line; these are shown in Table 4.8.

TABLE 4.8

PERCENTAGES OF "CLASS I" AND "NOT CLASS I" PUPILS ABOVE THE RED LINE

	Combined 'H'	Fourth Year 'O'	Total
"Class I" (Sample Number)	77% (35)	75% (8)	77% (43)
"Not Class I" (Sample Number)	53% (384)	63% (591)	59% (975)
Difference 95% Confidence Interval	24% {7%, 41%}	12% {0%, 64%}	18% {3%, 33%}

The very small number of Class I pupils in the Fourth Year 'O' grade sample gave rise to a very large Confidence interval. (The lower limit has been recorded as 0%, as a negative value is meaningless). The consequent uncertainty meant that this result could not be informative. The 'H' grade results, however, indicated that Class I pupils were more likely to be above the Red Line than other pupils. When the 'O' and 'H' grade results were combined, the same trend was observed.

These results were consistent with a relationship between recognition of Class I groups as units and their conceptualisation as such; however, because of the small number of Class I pupils they could be only suggestive, and would not be claimed as substantial evidence for such a relationship.

The second investigation made involved a comparison of the performance of forms. The Red Line categorisation was supplied for 22 'O' grade forms and 12 Fifth year 'H' forms. (It was also supplied for 5 Sixth year 'H' forms, but this number was too small for viability). It was proposed that the higher the mean Class I ratio of a form, the better the performance of that form should be, given the one-to-one correspondence specified above. The percentage of pupils in a form above the Red Line was taken as the measure of the form's performance in Chemistry, and forms within each year were ranked accordingly. They were also ranked in order of their mean Class I ratio Scores, and the rank correlation coefficient[75] was computed for each year. Formally, H_o: $r = 0$ was tested against H_1: $r \neq 0$. As no confidence intervals could be calculated for this statistic only the values of the coefficients and the probabilities, p, have been reported:

Fifth year forms: $r = 0.55$ ($.1 > p > .05$), ($N = 12$)

Fourth year forms: $r = 0.75$ ($p < .05$), ($N = 22$).

Although no confidence intervals could be quoted, the fact that the Fifth year forms' coefficient of .55 was on the border-line of significance indicated that a large measure of uncertainty was inherent in the results, due to the small sample sizes. The best interpretation of the results was that there was probably a measure of positive correlation between mean Class I Ratio and form performance in Chemistry, although it was not possible to specify accurately the magnitude of this relationship; that is to say, these results also were suggestive but inconclusive.

In summary, the correlation procedures reported in this Section gave conclusive results where Visual Scores were concerned. No evidence was found of a substantial relationship between the recognition of groups as units and Visual Ability. No evidence was found of a relationship between Visual Ability and performance in Chemistry.

The results pertaining to Class I ratio and performance in Chemistry were less informative. Because only a small number of pupils were assigned to Class I, and because the majority of the remaining pupils fell far short of the criterial performance, it was not possible to test, in the intended manner, the assumption that there was a one-to-one correspondence between recognition of Class I groups as units and the specified level of conceptual understanding of functional groups. The subsidiary investigations showed results that were consistent with such a relationship, but could not provide substantial support for it because of the small sample sizes involved.

4.6 The Visual Difficulties and the Conceptual Difficulties Hypotheses

In this Section, the status of the two hypotheses will be considered in the light of the results obtained.

4.61 The Visual Difficulties Hypothesis

The results to be considered in testing the hypothesis that pupils' difficulties were Visual in origin may be summarised as follows:

1. <u>Level of Visual Ability</u>

   (a) The mean Visual Score for all pupils of 7.7 (S.D. 1.6) would enable a pupil who recognized Class III groups as units to reproduce correctly a simple Ester formula, under test conditions (Section 4.32).

   (b) The overall mean and standard deviation were consistent with the ability to reproduce Patterns being limited by Short Term Memory Capacity, rather than Visual Ability itself (ibid.).

2. <u>Characteristics of Responses</u>

   (a) The universal tendency was for incorrect Pattern responses to be incomplete or inaccurate, but still sensible, representations of the original Items (Section 4.31). With one exception, the same tendency was observed in responses to Molecule Items (Section 4.43).

The responses given to the "box" Molecules, particularly by Fourth year pupils formed the exception to the tendency mentioned above. It will be recalled that all Items were drafted on a rectangular grid to avoid any suggestion of grouping. The normal classroom practice would be to draw such formulae in a different way - for example, $-CH_3$ groups would be drawn as $\underset{C}{\overset{H\ \ H\ \ H}{\diagdown\,|\,\diagup}}$ rather than $H-\overset{H}{\overset{|}{C}}-H$, and this would certainly make the box pattern of the formulae much less apparent. Therefore, this form of confusion might very well not arise in practice. However, the fact that a sizeable group of Fourth year pupils was confused by this characteristic would certainly suggest that teachers should take care to draw such formulae in a "non-box" format, particularly for junior pupils.

(b) The responses to the pattern Items categorized as having left-hand end confusion indicated that the addition of 1 point to their Difficulty Number adequately accounted for the characteristic (Section 4.31).

The fact that this point was awarded for a characteristic rather than a component of an Item indicated that such Patterns did cause confusion. However, there was no indication that Molecule Items caused such confusion, even where a long side chain was located near the left-hand end of a Molecule (e.g. Item 34, Figure 4.3). It could well be that this characteristic was confusing only within the context of the skeletal patterns used in the Pattern Test, in which case this source of confusion would not be of practical (or everyday) consequence.

In the everyday classroom situation, pupils would not be required to memorize formulae. The performance of pupils under test conditions, indicated by the results quoted above, suggested very strongly that their level of Visual Ability was such that normal Organic formulae should not cause confusion.

3. <u>The Effect of the Orientation of Groups</u>

(a) A specific examination of responses to Items containing left-handed and right-handed -O-H and $-C(=O)-O-H$ groups (Section 4.43) suggested strongly that "back-to-frontness" did not cause difficulties for pupils. This agreed with pupils' own comments (Section 2.3).

(b) The agreement between expected and observed numbers of correct responses for Items containing a functional group in several different orientations suggested strongly that in fact pupils were not confused by any inversion of groups.

4. <u>The Relation Between Visual Ability and Performance in Chemistry</u>

   (a) There was no indication that the differences in Mean Visual Score, going from the Fourth year 'O' grade sample to the Sixth year SYS sample were of material consequence; in fact, it seemed that there was probably not even a directional increase in Visual Score (Section 4.32). This must be contrasted with the fact that there is a clear improvement in Chemistry performance across the years sampled.

   (b) A correlation of Visual Score and performance in Chemistry (measured by position with respect to the Red Line) also failed to show any relationship between these variables, at least for the 'O' and 'H' grade samples.

Individually, these results failed to provide any evidence to support the Visual Difficulties Hypothesis. In total, they contra-indicated the Hypothesis that pupils' difficulties were Visual in origin.

4.62 The Conceptual Difficulties Hypothesis

The results pertinent to the testing of the Conceptual Difficulties Hypothesis, that pupils' difficulties were Conceptual in origin, were as follows:

1. <u>The Relation Between Visual Ability and the Ability to Recognize Groups as Units</u>

   (a) A correlation between Visual Score and Class I Ratio failed to show any evidence of a substantial relationship between Visual Ability and the ability to recognize groups as units (Section 4.5).

   (b) The overall mean Visual Score, and the failure to observe confusion in incorrect responses (with the exception mentioned above) suggested strongly that pupils would not be prevented from recognizing groups as units because of very low Visual Ability (Sections 4.31 and 4.32).

These two results suggested that the conditions (discussed in Section 2.52 and 3.3) which would have invalidated the use of the Combined Tests as a practical test of the Conceptual Difficulties Hypothesis did not obtain.

2. <u>The Recognition of Groups as Units - the Criterial Performance</u>

   (a) Only 5% of the total sample were assigned to Class I. The majority of pupils in each year were assigned to Class III (Section 4.42).

   (b) A comparison of predicted and observed numbers of correct responses for each Molecule Item indicated that it was valid to infer that pupils recognized as units the set of groups corresponding to their assigned Class (Section 4.41).

These results indicated that only 5% of the sample met the criterial performance of recognizing as units the common groups $-C(=O)-O-H$, $>C=O$, $-OH$, $-CH_3$, $>CH_2$.

3. <u>The Relation Between Group Recognition and Performance in Chemistry</u>

   (a) While the mean Class I Ratio Score for each sample of pupils was well below the criterial value of 1, there was a clear difference in means across years, with the greatest difference occurring between the 'O' and 'H' Grade groups. This differential response in means corresponded to the differences observed between the Class I-IV distributions of pupils in the different groups. The classification showed that there was no improvement between the 'H' samples and the SYS sample in terms of criterial performance; the difference between these groups lay in the distribution of pupils within the other Classes (Section 4.42).

(b) The level of performance shown by the majority meant that the proposed Correlation between Class I Ratio and performance in Chemistry could not be informative. It seemed that the percentage of "Class I" pupils above the Red Line tended to be greater than the percentage of "Non-Class I" pupils, and that form performance was correlated with mean Class I ratio, but these results were not conclusive.

Overall, although there was evidence of a general improvement in Class I Ratio across years, - and a clear difference between the proportions of '0' grade and more senior pupils recognizing the Class I groups as units - any attempt to make a more critical measure of the relationship between Class I Ratio and performance in Chemistry was frustrated by the characteristics of the samples.

We have already noted (Section 2.52) that pupils were not taught to recognize Class I groups as units. It would seem reasonable, therefore, to infer that pupils who showed such recognition did so because they conceptualised them as such. The point that could not be demonstrated clearly was that a lack of such conceptual understanding could necessarily be inferred from a failure to recognize these groups as units.

The demonstrated level of Visual Ability gave no grounds for supposing that the majority of pupils would have been <u>unable</u> to recognize the Class I groups as units (i.e. chunk them), had they conceptualised them as such. Indeed, only some 10% of the total sample recognized no groups as units.

Altogether, approximately 84% of pupils recognized either the Class II or Class III groups as units. It is very difficult to see why these pupils should perceive the $\overset{O}{C}$-O-H group, for example, as a collection of somewhat arbitrary 'bits' (C, $\overset{O}{\|}$, -OH for Class II pupils, or C, $\|$, O, -OH for Class III pupils) if in fact they conceptualised this group as a single Chemical entity. It would seem more likely that these pupils recognized as units various formula 'bits' that were familiar to them. The bits $\overset{O}{\|}$, or $\|$ and O, in isolation, are not chemically informative. Given the presence of a $\overset{O}{\|}$ unit, the <u>chemist</u> needs to know whether the molecule contains a $>C{=}O$, $-C\overset{O}{/\!/}-OH$, or $-C\overset{O}{/\!/}-O-$ group. The fact that pupils recognized the important $-C\overset{O}{/\!/}-OH$ and $>C\overset{O}{/\!/}$ groups as collections of bits rather than single units, suggested strongly that a lack of chemical conceptual understanding underlay their recognition behaviour.

Considering these arguments, along with the results summarised above, we would suggest that it was probably unlikely that the Class II, III and IV pupils had attained the specified level of conceptual understanding of Functional Groups.

4. <u>The Recognition of Functional Groups First</u>

> There was no indication that pupils reproduced a functional group correctly in an otherwise incorrect Molecule Item response (Section 4.4). Thus, there was no evidence to suggest that pupils instinctively noted the functional group(s) in a formula first.

This latter result, and the very small proportion of pupils recognising the Class I groups as units, both supported the Conceptual Difficulties Hypothesis.

4.7 <u>Conclusion</u>

Overall, the Combined Tests showed no evidence in support of the Visual Difficulties Hypothesis, but did show evidence consistent with the Conceptual Difficulties Hypothesis. It seemed possible that only a very small proportion of the large, representative sample tested had attained the specified level of conceptual understanding of Functional Groups.

It was therefore decided to concentrate further investigations on the Conceptual Difficulties Hypothesis, and in particular on obtaining answers to two questions:

(i) Was there a widespread failure to attain the specified level of conceptual understanding?

(ii) Was the lack of this level of conceptual understanding relatable to the mistakes pupils made in the tasks they were normally required to perform in Organic Chemistry?

While it could have been possible to administer the Combined Tests to a different and much larger sample of SYS pupils, and at the same time to measure their performance specifically in Organic Chemistry, this approach was rejected. A larger sample would have given a smaller confidence interval, and the use of a specific, Organic Test would have probably increased the accuracy of the comparison between Class I Ratio and performance in Chemistry. However, the distribution of SYS pupils in the sample tested across Classes I-IV indicated that a correlation procedure had no a priori guarantee of success, even with this more restricted sample.

It seemed possible that definite answers could be obtained to both questions by using a different approach. This involved an experimental investigation of the strategies employed by pupils in Organic Chemistry tasks. This investigation, which in fact implements Stage 5 of the Experimental Design (Section 2.53), is described in the next Chapter.

APPENDIX 4.1

Copy of Instruction Sheet for Participating Teachers

Dear

I recently wrote asking for your co-operation in giving a test related to Organic Chemistry. Thank you very much for your offer of help.

I have sent you the test materials now as I felt this would give you more freedom to choose a convenient time for giving the test.

When you have completed the test, please return the answer sheets to me - there is no real need to return test-cards or transparencies. I realise that there may be problems in returning large sets of answer sheets by post - if so, please let me know, and I will try to arrange a collection date.

I hope that the enclosed instructions are quite clear but if you have any doubts, please let me know.

(1) TEST MATERIALS

You should have received -

1 or 2 sets of pattern cards
1 or 2 sets of formulae cards, and/or
1 set of pattern transparencies
1 set of formulae transparencies
and sufficient answer sheets to give 1 to each pupil.

(2) ADDITIONAL MATERIALS

Pupils will need a pencil or pen. You will need a stop-clock.

(3) THE TESTS

The PATTERNS are shown first;
the FORMULAE are shown second.

Each pattern (or formula) is shown for 10 seconds. A pause of about 18 seconds is allowed for pupils to draw what they have seen.

For some of the simpler patterns, a much smaller drawing time can be allowed - I went on to another pattern if it was quite clear that everyone had finished, but no more than 20 seconds should be given for the most complex patterns. At the end of the "reponse time", it is best to say "ready", "next one", or something similar, to return pupils' attention to the new card.

When students have filled in one page, they should be given sufficient time to turn to the next page - and wriggle a bit!

The answer sheets have been arranged so that the pattern answers are recorded on the fronts of the pages, going across the page. When all the patterns have been shown, pupils should turn the whole answer block over, and record answer formulae on the backs of the pages.

The two tests will take approximately 18 minutes each, and may I remind you that a rest between the two is very necessary.

(4) USING TEST CARDS

To be shown in the order provided. They are numbered on the back in the top right hand corner.

(5) USING TRANSPARENCIES

The best method of using these that we have found is to cover a transparency with two sheets of paper. Pull the lower sheet down to expose a pattern, then at the end of the 10 seconds, pull the upper sheet down to cover it again. (You'll probably find a trial run helpful).

(6) GIVING THE TESTS

The first pattern is just for demonstration - pupils don't copy it. It makes sure they can see the patterns easily, and know what to look for.

Show the first pattern and say something like the following - "You are going to take part in an experiment which is being done in many Scottish schools. You are going to be shown a lot of patterns like this. You will be shown a pattern for 10 seconds. When I cover it up, draw what you have seen on your answer sheet. Put your answers across the sheet. You will have only about 20 seconds to draw the pattern, so don't try to be neat. Don't use a ruler. The dots and shapes you see are shaded in - that is just to make them easier to see. Don't try to shade them in when you draw them".

When you start the formula test, just say they will see structural formulae this time, but to answer just as before. There is no demonstration formula.

At the end of the test, pupils might like to know that the experiment will (hopefully!) tell us why pupils have found difficulty with topics such as condensation, esterification, etc.

(7) FILLING IN "YEAR, FORM, CODE"

YEAR: Could pupils put the year of chemistry they study

FORM: 5A, 4C, or whatever

CODE: Your school code is , so please put this number in first.

I would like to correlate part of my results with some measure of pupils' "chemical ability", and I wonder if you could help me here by letting me know the position of each pupil in the order of merit list. I thought this could be done by getting each pupil to write his order number after the school code. If you think this might give pupils information you don't want them to have, perhaps you could give them a code number or letter, and send me a list indicating positions. Could you also indicate where the red line is drawn - either on the appropriate answer sheet or separately. I will be very grateful for the information in whatever form you find easier.

May I thank you once again for your help. I will be very happy to send you the results of this investigation as soon as possible.

Yours sincerely,

(Mrs.) Natalie C. Kellett

APPENDIX 4.2

Details of the Sample Used in the Combined Tests

TABLE 4.A2.1

DESCRIPTION OF PARTICIPATING SCHOOLS

School Identity Number	Description of School
1	Large Urban Comprehensive Co-educational
3	Large Urban Comprehensive Co-educational
4	Urban Comprehensive Co-educational
5	Roman Catholic Urban Comprehensive Co-educational
6	Urban Comprehensive Co-educational
7	Rural Private Selective Co-educational
8	Large Urban Comprehensive Co-educational
9	Rural Private Selective Co-educational
10	Urban Comprehensive Co-educational
11	Urban Comprehensive Co-educational
13	Urban Comprehensive Co-educational
14	Large Urban Selective Boys
15	Large Urban Selective Boys
17	Large Urban Selective Boys
18	Urban Comprehensive Co-educational
19	Urban Comprehensive Co-educational

TABLE 4.A2.2

A BREAK-DOWN OF THE SAMPLE USED FOR THE COMBINED TESTS

	Sixth Year SYS		Sixth Year 'H'		Fifth Year 'H'		Fourth Year 'O'	
School	Number Forms	Total Pupils	Number Forms	Total Pupils	Number Forms	Total Pupils	Number Forms	Total Pupils
1	1	10	1	16	1	25	2	117
3	1	10	-	-	1	20	1	21
4	1	7	-	-	3	39	3	35
5	-	-	-	-	-	-	3	63
6	1	9	-	-	1	43	1	46
7	-	-	-	-	-	-	4	53
8	1	4	1	9	1	10	1	11
9	-	-	-	-	2	32	3	43
10	1	9	-	-	1	15	2	30
11	-	-	-	-	1	26	-	-
13	-	-	1	9	2	32	1	12
14	-	-	1	15	1	34	1	66
15	-	-	1	14	1	113	1	117
17	-	-	2	35	1	20	-	-
18	-	-	1	14	-	-	2	28
19	-	-	1	7	1	18	3	55
Totals	6	49	9	119	17	427	28	697

APPENDIX 4.3

Derivation of a Formula for the Maximum Width of a Confidence Interval

The 95% Confidence Interval about the difference between two proportions is approximately

$$(p_1 - p_2) \pm 1.96 \sqrt{p(1 - p)(1/N_1 + 1/N_2)},$$

(where p_1, p_2, are the proportions in the samples of size N_1, N_2),

and $$p = \frac{N_1 p_1 + N_2 p_2}{N_1 + N_2}$$ is the mean proportion.

The width of the Confidence Interval, W, is given by the second term in the above expression;

$$W = 1.96 \sqrt{p(1 - p)(1/N_1 + 1/N_2)}.$$

The Confidence Interval will not capture 0 if

$$|p_1 - p_2| > W.$$

Although W depends on the values of p_1 and p_2, as well as on the sample sizes, it has a maximum value for any fixed N_1 and N_2, that depends only on the sample sizes.

For fixed N_1 and N_2,

$$W = C \sqrt{p(1 - p)} \qquad C = 1.96 \sqrt{1/N_1 + 1/N_2}$$

$$\frac{dW}{dp} = \frac{\frac{1}{2}C(1 - 2p)}{\sqrt{p(1 - p)}}$$

$$= 0 \text{ when } p = \tfrac{1}{2}.$$

When $p = \frac{1}{2}$, W takes on its maximum value,

$$W_{max} = \frac{1.96}{2} \sqrt{1/N_1 + 1/N_2} \ .$$

Let $$x = \frac{N_1}{N_2}$$

then $$W_{max} = \frac{1.96}{2\sqrt{N_2}} \sqrt{1 + 1/x} \ .$$

In the general case, it is convenient to define a new variable,

$$w = \frac{1.96}{2} \sqrt{1 + 1/x} \ .$$

Values of w may be recorded (by tabulation or graphically) for

$.01 > x > 1$.

If $$|p_1 - p_2| > W_{max},$$

i.e. $$> \frac{w}{\sqrt{N_2}}$$

then the Confidence Interval about $(p_1 - p_2)$ will not capture 0, whatever the values of p_1 and p_2.

If percentages rather than proportions are used, the required difference is

$$\frac{100w}{\sqrt{N_2}}$$

In the present case, for the combined sample, $N_1 = N_2 = 1292$, and $x = 1$. Therefore, the value of W_{max} is

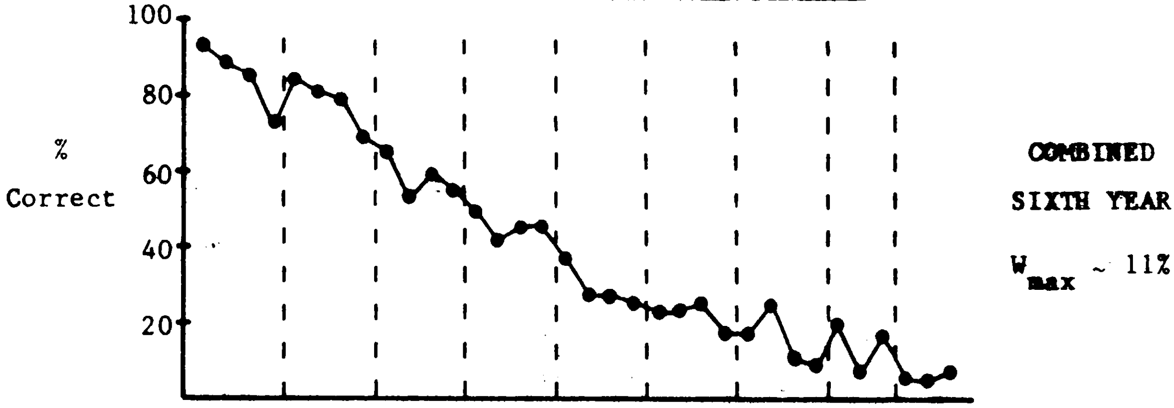
$$W_{max} = \frac{100.1.96}{\sqrt{2} \ . \ \sqrt{1292}}$$

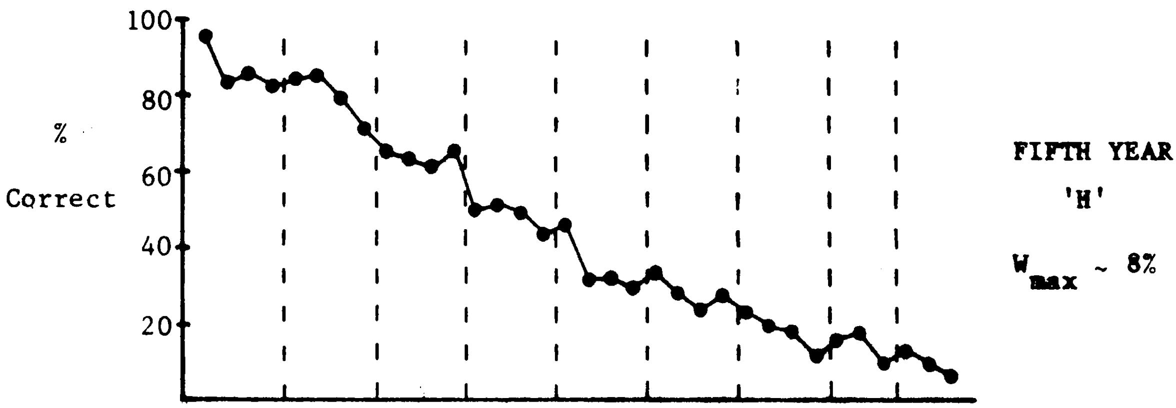
$$= 3.87\%$$

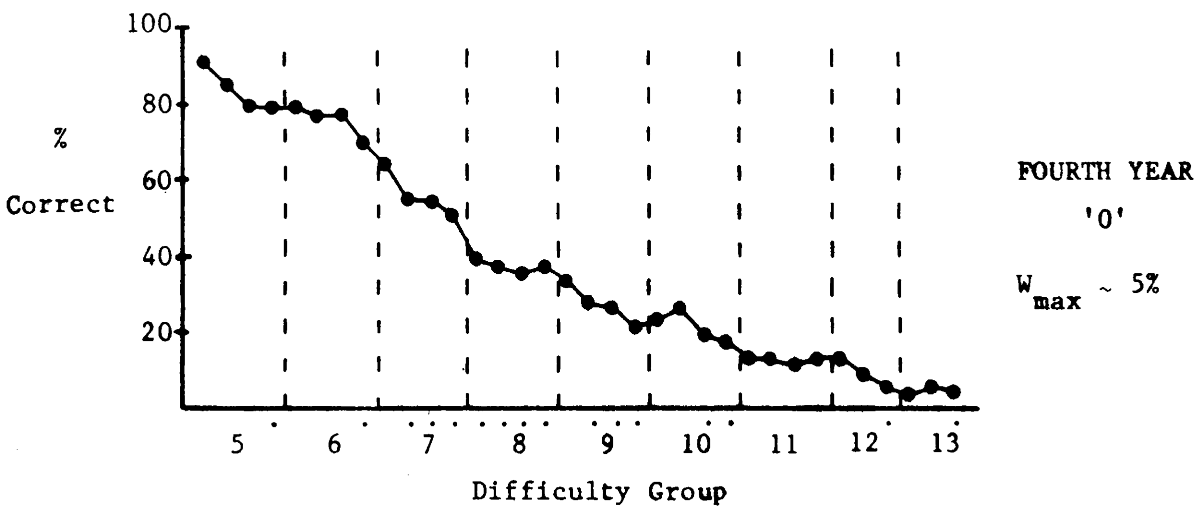
$$\simeq 4\%.$$

APPENDIX 4.4

Percentage of Pupils Within Each Year Giving Correct Responses for each Pattern Test Item







APPENDIX 4.5

Expected Numbers of Correct Responses for Molecule Items: Derivation of the Computational Formula

For simplicity, we will derive the formula for determining the number of correct responses expected from pupils assigned to one particular Class - the ith Class.

Consider a Molecule Item having Difficulty Number n when the Class i groups are scored as units. The number of correct responses expected from Class i pupils, N_i, would be in principle,

$$N_i = I_n^i, \text{ the number of persons in Class i having Visual Score} > n.$$

In practice, two correction terms were introduced, to obtain a more accurate prediction model. In each case, the simplest possible correction factor was used, to avoid building prejudice into the model.

1. A student could be awarded a Group "correct" if he had only 3 of its 4 Items correct. Therefore, we could not assume that a pupil having, say, Visual Score 6 should reproduce correctly <u>every</u> Item whose "i" Difficulty was less than or equal to 6. The minimum response rate for a "correct" Group was 3/4, and the maximum 1; the Expectation Value for correct responses was therefore taken to be 7/8.
   The first correction was, therefore, to set
   $$N_i = (7/8)\ I_n^i\ .$$
2. If a pupil had Visual Score 6.5, say, he must have reproduced correctly at least two Items of Difficulty greater than 6. Observation of pupils' response sequences had shown that a "6.5"

pupil would not necessarily have his two "extra" correct Pattern Items in Group 7.

To allow for the non-integer Visual Scores in the simplest way possible, we specified an Expected Assignment of these "extra" Items for the three fractional increments ¼, ½ and 3/4;

m + ¼: 1 Item from the 4 Items in Difficulty Group m + 1 correct.
Expectation Value for an "m + 1" Item = .25.

m + ½: 1 Item correct in Difficulty Group m + 1.
1 Item correct in Difficulty Group m + 2.
Expectation Values: "m + 1" Items = .25
"m + 2" Items = .25

m + 3/4: 2 Items correct in Difficulty Group m + 1.
1 Item correct in Difficulty Group m + 2.
Expectation Values: "m + 1" Items = .5
"m + 2" Items = .25

(As the last two Pattern Groups contained only 3 Items, Visual Scores with fractional increments of .33, .67, etc. were possible. In these cases, the number of "extra" Items was used to allocate a pupil to the ¼, ½ or 3/4 category).

Using this specification, we could compute the number of expected correct responses for the Item of Difficulty n from the pupils having Visual Score n - 1 or n - 2, plus a fractional increment.

Let I^i_{1n} = No. pupils having Visual Score (n - 1) + ¼

I^i_{2n} = No. pupils having Visual Score (n - 1) + ½

I^i_{3n} = No. pupils having Visual Score (n - 1) + 3/4

I^i_{4n} = No. pupils having Visual Score (n - 2) + ½

I^i_{5n} = No. pupils having Visual Score (n - 2) + 3/4.

Then the contribution from these pupils to N_i was found by summing

$$(\text{Expectation Value})_k \times I^i_{kn}, \quad (k = 1,5),$$

giving a total contribution of

$$.25\left[\sum_{k=1}^{5} I^i_{kn} + I^i_{3n}\right]$$

Therefore, the final computational formula for pupils from the ith Class was:

$$N_i = (7/8)\, I^i_n + .25\left[\sum_{k=1}^{5} I^i_{kn} + I^i_{3n}\right].$$

To obtain the number of expected correct responses from pupils in all Classes, we simply summed the N_i's, to give N_E;

$$N_E = \sum_{i=1}^{4} N_i .$$

Although this formula may appear cumbersome, it was very simple to use in practice. Starting from the two-way classification table (Table 4.A5.1), a cumulative count enabled I^i_n (and hence $7/8\ I^i_n$) to be determined for each integer Difficulty Number, for each Class. Next, for each integer Difficulty Number, the non-integer contribution from the two preceding Visual Score groups were computed and hence the four N_i's for each possible integer Difficulty Score Number were determined. To obtain the total number of expected correct responses for any Molecule Item, we had only to add the four N_i's corresponding to the Difficulty Numbers awarded to the Item for each Class.

One further procedure was used; where a Molecule Item had an (integer + ½) Difficulty Number, the expected number was determined by interpolation. The Difficulty Numbers assigned to each Item are shown in Table 4.A5.2, together with the expected and observed numbers of correct responses, and the value of χ^2_1, for each Molecule Item (for the combined sample).

TABLE 4.A5.1

CLASSIFICATION OF PUPILS BY VISUAL SCORE AND CLASS

Visual		Sixth Year SYS				Sixth Year 'H'				Fifth Year 'H'				Fourth Year '0'			
Score		I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
4	.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	.25	-	-	-	-	-	-	-	-	-	-	1	-	-	-	1	-
	.5	-	-	-	-	-	1	-	-	-	-	1	-	-	-	4	-
	.75	-	-	1	-	-	1	-	-	2	1	2	-	-	1	9	-
5	.0	-	1	1	-	-	-	2	-	-	3	3	-	-	1	14	-
	.25	-	-	-	-	1	-	-	-	-	3	2	-	-	2	12	1
	.5	-	1	-	-	-	-	3	-	2	1	5	-	-	4	23	4
	.75	-	-	1	-	1	-	6	-	2	3	9	-	-	-	25	-
6	.0	-	-	-	-	1	1	6	-	1	3	9	-	-	1	29	1
	.25	-	-	-	-	-	1	7	-	-	6	7	1	-	-	39	1
	.5	-	1	1	-	1	2	8	-	4	4	12	1	1	1	33	5
	.75	-	1	-	-	-	-	3	1	2	5	18	2	1	1	39	6
7	.0	-	-	3	-	1	-	6	-	1	2	22	1	-	1	35	3
	.25	-	1	1	-	3	1	5	-	2	2	18	1	3	-	45	7
	.5	-	-	-	-	-	-	12	-	5	4	21	1	1	3	30	9
	.75	1	-	3	-	1	-	8	-	1	5	19	2	-	3	36	10
8	.0	2	1	2	-	-	2	3	-	5	6	13	1	-	3	27	4
	.25	-	2	3	-	1	-	4	-	3	3	18	1	1	-	32	9
	.5	-	1	2	-	-	1	4	-	1	8	14	-	1	2	24	5
	.75	-	1	-	-	-	-	3	1	2	5	14	1	-	1	18	6
9	.0	-	-	2	-	1	-	1	-	-	2	11	2	-	1	19	6
	.25	-	-	2	-	-	-	1	1	-	-	12	1	-	-	17	8
	.5	-	-	2	-	-	1	-	1	-	1	16	1	-	-	10	7
	.75	-	-	3	-	-	-	2	-	1	-	6	1	-	-	7	1
10	.0	-	-	2	-	-	-	-	-	-	1	8	4	-	-	11	4
	.25	-	-	-	-	-	-	3	-	-	-	4	-	-	-	1	2
	.5	-	-	1	-	-	-	2	-	-	-	5	-	-	-	5	-
	.75	1	-	-	-	-	-	-	-	-	-	2	3	-	-	3	2
11	.0	-	-	-	-	-	-	1	-	1	-	3	2	-	-	3	1
	.25	-	-	1	-	-	-	1	-	1	-	3	-	-	-	-	1
	.5	-	-	1	-	-	-	-	-	-	-	2	4	-	-	7	1
	.75	-	-	1	-	-	-	2	-	-	-	3	-	-	-	-	-
12	.0	-	-	-	-	-	-	-	-	-	-	2	-	-	-	1	1
	.33	-	-	1	-	-	-	-	-	1	1	-	-	-	-	-	-
	.67	-	-	1	-	-	-	-	-	-	-	1	1	-	-	-	-
13	.0	-	-	-	-	-	-	-	-	-	-	4	-	-	-	-	-
Totals		4	10	35	0	11	11	93	4	37	69	290	31	8	25	559	105

TABLE 4.A5.2

EXPECTED AND OBSERVED NUMBERS OF CORRECT RESPONSES (COMBINED SAMPLE)

Item Number	Difficulty Number of Each Item/Class I	II	III	IV	Expected Number Correct	Observed Number Correct	Chi-Squared Value
1	5	5	5	6	1122.88	1148	2.294
2	1½	3½	3½	6½	1193.19	1176	1.500
3	4	4	4	6	1127.00	1121	0.123
4	4	4	4	6½	1122.63	1107	0.798
5	4	5	6	7	1043.25	1068	1.568
6	6	6	6	7	1025.80	881	*
7	5	5	6	7	1042.00	1060	0.826
8	5	5	5	7½	1094.63	1083	0.395
9	5	6	7	8½	847.19	845	0.008
10	5½	5½	5½	8½	1030.19	999	2.233
11	3	4½	6½	8	953.76	942	0.274
12	6½	6½	6½	8½	902.75	868	2.167
13	1½	5½	8	9	631.77	631	0.001
14	6½	6½	6½	9	890.20	574	*
15	4	6	7	9½	820.69	803	0.518
16	7½	7½	7½	9½	666.63	650	0.428
17	6	7	7½	10	676.50	710	1.746
18	3	5	7½	10	718.17	729	0.184
19	3	7	9	10	402.38	406	0.024
20	3	6	9½	10	356.94	347	0.193
21	5	6	7½	11	682.38	720	2.207
22	5½	7½	8½	11½	452.96	415	2.500
23	6½	7½	9	11	363.07	356	0.096
24	7½	7½	7½	11	635.00	443	*
25	9	9	9	12	279.79	305	1.405
26	9	9½	9½	12	210.41	206	0.056
27	9½	9½	9½	12½	206.62	198	0.218
28	8	8½	9	12	313.54	319	0.062
29	9½	9½	9½	13	205.38	225	1.073
30	7	9	9½	13	242.56	255	0.385
31	6	8½	10	13	201.56	206	0.057
32	6	9	10	13	186.30	170	0.864
33	10	10	11	14	76.39	246	*
34	8	9	10	14	162.59	183	1.392
35	8	10	10½	14	116.27	125	0.348
36	10	10½	10½	15	94.60	96	0.011
37	9	10	10½	15	101.93	96	0.192
38	10	11	11½	15	42.25	73	8.590
							34.75 (34°F)

* indicates that a value of χ_1^2 was not calculated

CHAPTER 5

A Further Examination of the Conceptual Difficulties Hypothesis: The Grid Test Experiment

For the sake of clarity, this second phase of the testing of the Conceptual Difficulties Hypothesis will be referred to as the 'Grid Test' Experiment (so named because of the format of the major test materials used in the investigation).

In broad terms, the investigation described in this Chapter sought to determine the strategy (or strategies) employed by pupils in answering three types of question:

(1) Is compound A in the same family as compound B (given their formulae).

(2) Will the reactions of compound A include the characteristic reactions of compound B (given their formulae).

(3) In what ways are reactions A and B the same or different (given the equations).

The way in which knowledge of pupils' strategies could be used in conjunction with the information obtained from the Combined Tests to answer the questions proposed in the concluding section of Chapter 4 becomes apparent when we consider how questions of type 1-3 may be answered. Basically, there are two possibilities. A and B may be familiar objects, with knowledge of them including facts that specify the same/different relationship. Alternatively, A and B may not both be familiar objects. In the latter case, if a correct answer is to be obtained, there will be two requirements:

(a) A relevant set of necessary and sufficient criteria for judging sameness/difference must be possessed, and

(b) The characteristic properties of A and B, on which the judgement criteria are to operate, must be known and correctly identifiable.

Consider, for example, a student who knew that ethanoic acid was a carboxylic acid, and could recognize its formula, and also knew that butanoic acid was a carboxylic acid and could recognize its formula. If he were shown the formulae of these two compounds he would be able to state that the compounds belonged to the same family, on the basis of his knowledge of these particular familiar objects. However, if this student were shown two completely unfamiliar formulae, (or indeed one familiar and one unfamiliar formula), he would need to know that the salient characteristic property of each compound was its functional group, and would have to be able to distinguish both functional groups correctly. A relevant set of necessary and sufficient criteria would be:

(i) Are the functional groups the same? (The 'necessary' condition for 'sameness'). A decision based on this criterion could be modified by

(ii) Will the environment of the functional group, in either case, modify its chemical behaviour to an appreciable extent? (The 'and sufficient' criterion).

In answering any of the question types 1-3 listed above concerning unfamiliar compounds, the strategies a pupil employed would be his choice of the salient characteristic property, his method of identifying or distinguishing it, and his choice of judgemental criteria. In each of the three cases, the strategies adopted by a pupil would depend upon, and therefore reflect, his level of conceptual understanding of functional groups.

In line with the work already reported, we were particularly interested in determining whether pupils treated functional groups as units, or as collections of "bits" - if in fact they chose functional groups as the characteristic property. Furthermore, observation of the strategies employed could show whether a failure to treat functional groups as units, and behaviour determining units, (i.e. a lack of attainment of the specified level of conceptual understanding) was directly relatable to the mistakes pupils made in answering these questions, which are typical and important tasks required of pupils studying the Scottish syllabus.

The Experimental Design, and the test materials produced, will be described in Section 5.1. The results will be presented and discussed in Section 5.2. These results posed a very interesting question concerning pupils' choice of Condensation, Hydrolysis and Esterification topics as an area of difficulty, and led to the proposal of an hypothesis concerning the adjudged difficulty of Chemistry topics. This hypothesis forms the subject of the final section of this chapter.

5.1 Experimental Design

5.11 The Selection of the Test Sample

The pupils who participated in this experiment were all fifth years, who had completed the 'H' Grade Organic Chemistry course a short time before the administration of the tests. This cohort of pupils had had at least two years' experience of Organic Chemistry, and it was therefore reasonable to hope that they had developed strategies for answering the types of questions of interest. Furthermore, many schools had large numbers of fifth year Chemistry students, which obviated any problems of sample size. (SYS pupils would have formed an interesting participatory group for this experiment, but in their case, because the number of SYS pupils per school is generally small, we would have expected sample size problems).

Pupils were drawn from the (representative) set of schools that had participated in the Combined Tests experiment. Each school was asked to provide a representative group of 15 to 20 pupils. There was, of course, no certainty that all these pupils - or even a majority of them - would have also participated in the Combined Tests experiment during the previous year. However, the generalizability of the results of that experiment allowed the assumption to be made that a large percentage of these pupils would not have recognized as units all the important functional groups (the Class I set of groups).

5.12 The Experimental Procedures

The investigation was planned as a three stage experiment, in which the three stages were run concurrently.

Stage (i) The Grid Test

The Grid Test consisted of a series of four exercises in which pupils were asked to determine which compounds from a given series of formulae were in the same family as an exampled compound, or showed the same characteristic reaction as an exampled compound - that is, pupils were required to give responses to the first two of the question types listed at the beginning of the Chapter.

As no statistical analyses were to be applied to the results of the Grid Test, there was no formal lower limit on the sample size. However, for the results to be generalizable, the sample had to be representative; this requirement was met by drawing pupils from a representative set of schools.

The Grid Test was administered to all pupils participating in the experiment. The selection of items for this test, and its associated scoring system are described in detail in Section 5.131.

Stage (ii) The Interview

Interviewing offered the possibility of collecting information that would otherwise have been unobtainable; for example, the possibility of getting pupils to amplify or elaborate upon a response, and the possibility of getting them to respond orally to questions in cases where they could have found detailed written responses a formidable task. Accordingly, arrangements were made to interview pupils from two schools immediately after they had completed the Grid Test.

First, each pupil was to be asked to go through one of his Grid Test exercises, and explain how he had rejected the items he considered to be "different" from the exampled compound. These comments were to act as a check upon, and a guide to, the interpretation of the Grid Test results.

Secondly, pupils were to be asked, for each of the carboxylic acid, ester and alcohol families, "If someone asked you what a (... family ...) was, what would you say?" Although pupils could give rote learned responses to such questions, we hoped to obtain an indication of whether or not they would define a family in terms of a functional group.

In the third interview question, pupils would be shown the three formulae NaOH, H_2O and CH_3OH, and asked - initially without prompting - what was the same and what was different about these compounds. If pupils made no comment about their behaviour, they would be asked, more specifically, if they would not expect some similarity of behaviour as each compound contained an -OH group. This was felt to be a difficult question for 'H' Grade pupils (although it would have been considered to some extent in class work), and was included in the interview because it seemed possible that many pupils would provide sparse and uninformative written responses.

The two final questions, which were identical in form, were included in the interview for the same reasons. In both cases, a pupil would be shown a pair of equations and asked if there were anything the same about the reactions, and if there were anything different about them. These two questions were intended to probe the relationship pupils saw between functional groups and chemical reaction.

Stage (iii) Reduction of the Effect of the Lack of Conceptual Understanding

While the first two stages of this experiment were designed to look for evidence of the lack of conceptual understanding of functional groups and evidence of a direct relationship between such a lack and the mistakes pupils made, the third stage involved a rather different approach. If pupils made mistakes because they had not attained the specified level of conceptual understanding, then one would expect their performance to improve if the effect of that lack of conceptual understanding could be removed or lessened.

Therefore, in the third phase, pupils from two schools would participate in the Grid Test in the ordinary way, and then during the ensuing week would use some specially prepared learning materials, designed to encourage pupils to choose the functional group as the characteristic property, and to treat functional groups as units, not collections of bits. Finally, these pupils would be asked to sit the Grid Test again.

The learning materials took the form of a card game, which is described in Section 5.133. While it is possible that the use of this card game could assist a pupil in attaining the specified level of conceptual understanding, such a transition may take a considerable time, as we have already emphasized in Chapter 2 (p. 47 et seq.). It follows that we could not expect that within just one week a change would necessarily occur in conceptual understanding sufficient, for example, to enable a pupil to recognize the Class I groups as units when he had not formerly acted in this way. It was for this reason that we have described the desired outcomes of using the card games in terms of the adoption of certain strategies (which, of course, one would expect to be used by a person who had attained the specified levels of conceptual understanding), and of a consequent removal of, or reduction in, the effect of the lack of conceptual understanding, rather than in terms of a direct change in conceptual understanding.

The use of the Grid Test as a pre- and post-test enabled a measurement of any change in performance that occurred after the week's use of the card game, both in terms of overall performance in each exercise and in terms of the strategies pupils used.

The number of pupils required for this phase of the experiment was determined by the condition that the power of a t-test of dependent means should be at least 0.8 (at $\alpha = 0.05$) for observing a medium to large difference between the pre- and post-test means. Cohen's tables[76] give a minimum number of 26 for this condition, and therefore two schools were asked to participate.

The Grid Test was suitable for use as a pre- and post-test because, as the description of the test format will show, there was only a very small likelihood of pupils' performance in the post-test being influenced by recall of this responses in the pre-test. The pupils who participated in this phase of the experiment were not told that there was any connection between the pre-test and the use of the card game;

nor were they given any prior notification that they would be required to sit a post-test.

5.13 The Test Materials

5.131 The Grid Tests

Each of the four exercises contained in the Grid Test was presented on a separate test sheet as a grid of sixteen numbered cells (see Tables 5.3-5.6, pp. 178-81).

The first three exercises were concerned with the selection of compounds "in the same family" as an exampled compound; the exemplar compounds were, respectively, a carboxylic acid, an ester, and an alcohol. In each case, the exemplar formula was located in cell 1, and the test formulae filled the other 15 cells. Pupils were required to answer on each test sheet, giving the name of the substance in cell 1 (or at least its family) and recording the number of each cell that contained a compound belonging to the same family as the compound in cell 1. Pupils were asked to name the exemplar so that they would study it carefully. A family exemplar formula was given (in preference to using the question "Which of the following are esters" - for example) to avoid contamination of the results due to a failure to recall, or an incorrect recollection of, the functional group associated with a family name.

The response format used - simply listing a series of numbers on a test sheet - allowed each exercise to be completed quickly. Because of this, and because pupils did not have to write any of the formulae, there seemed little chance that pupils, sitting the Grid Test as a pre-test for Stage (iii), would memorize any of the formulae they selected.

Two techniques were used in selecting the formulae used as test items. To determine whether pupils could distinguish the functional group correctly, and what their conception of the functional group was, a variation in functional group was used. To determine the extent to which pupils used the criterion "Do these compounds have the same functional group" as a determiner of family membership, a variation in environment was used.

Of course, the difficulty in selecting test items was to use sufficient variation in functional group and environment to show up the use of arbitrary or incorrect strategies without providing so much variation that the test would be outwith the competence of pupils, even if they had obtained a perfectly acceptable "'H' Grade" level of conceptual understanding.

Four variations in functional group were used, namely:

(i) A non-structural representation (e.g. -COOH for $-\overset{O}{\overset{/\!\!/}{C}}-O-H$). It was realized that pupils from different schools could have had different degrees of acquaintance with the use of non-structural representations; however, earlier discussions with teachers had suggested that structural formulae were used almost exclusively at this level, and therefore it seemed that differences in school experience would not be great. This variation was used to determine the extent to which pupils looked for a particular representation of a functional group, rather than for the presence of the group in the formula, however it was represented.

(ii) Different orientations of the functional groups.

(iii) 'Trick' functional groups - for instance, the inclusion of a molecule in the carboxylic acid grid that contained the groups $>\overset{O}{\overset{/\!\!/}{C}}$ and -O-H, on different carbon atoms.

(iv) Instances of other functional groups.

'H' Grade pupils would have had some knowledge of the effect of variation in the environment on the characteristic behaviour of functional groups, and this provided a guideline for the type of variation to be used in the Grid Test. For example, pupils would have studied the change in physical properties with increase in molecular mass in an homologous series. They would also be required to classify alcohols as primary, secondary or tertiary - an environmental classification. However, they would not have considered why this variation gave rise to three sub-classes of a family, whereas the effect of the environmental variation $-C^{\nearrow O}-H$ (considered as a limiting case of $-C^{\nearrow O}-R$) was considered sufficiently great to classify $-C^{\nearrow O}-H$ as a separate functional group, rather than the carbonyl group in a highly modifying environment.

With this background in mind, the following variations in environment were used:

(i) Replacing the carbon within a functional group, or adjacent to it, with another element.

(ii) Replacing one or more hydrogens with a halogen.

(iii) Using a derivative.

(iv) Inclusion of a double bond within the chain.

(v) Inclusion of another functional group.

It was realized that the last variation could represent a difficult task for these pupils, as they could quite reasonably be uncertain of whether or not such a 'difference' would outweigh the 'sameness' of the common functional group. It should be emphasized that in analyzing the results of this test, we would be interested in the relative effect that the different variations had on pupils' patterns of response, rather than on whether particular answers were 'right' or 'wrong'.

The Test Items were panelled by a group of experienced Chemistry teachers; this process indicated that two items were of doubtful value. One item was an amino acid (in the first, carboxylic acid, grid); some pupils had studied these compounds, and could have decided that this item was not a carboxylic acid, because it had a different name. The second item - one of the formulae in the Alcohol exercise - was judged to be outwith pupils' competence. Unfortunately, as some schools had already received the test materials, these items could not be replaced, so the alcohol formula (Item 9, Table 5.5) was excluded from the analysis of results, while the amino acid was included, but pupils were scored 'correct' for this item, whatever decision they made about its family membership.

The procedure to be adopted in reducing the ambiguity of the question concerning the alcohol family was also discussed with these teachers. Obviously, the exemplar formula had to be a primary, secondary or tertiary alcohol, but we required pupils to make only a general classification (i.e. 'alcohol' or 'not alcohol') of Test Items. It was agreed that most pupils would assume that a question implied the alcohol family generally, particularly where a primary alcohol had been used for the exemplar, unless specific mention were made of the sub-families. Therefore the wording of the question used for the first two grid sheets was judged to be appropriate for the alcohol sheet also. However, two additional measures were taken. Teachers who administered this Test were alerted to the possibility of this ambiguity, and instructed to inform any pupils who asked that the question referred to the "general" family membership. Secondly, any pupil who showed evidence of selecting only primary alcohols (the sub-class of the exemplar) would be judged to have used a perfectly acceptable strategy.

The fourth exercise was concerned with the identification of compounds that showed the same characteristic reaction(s) as a specified compound. While we expected to observe the use of the strategies adopted for the first three exercises, this exercise was included particularly to obtain information about the extent to which pupils related the chemical behaviour of a compound to its functional group(s), or considered it to be a property of the molecule as a whole. A slightly different procedure was adopted, to achieve this purpose. Two exemplar formulae were given - a carboxylic acid and an alcohol (Items 13 and 12, Table 5.6) - and two hydroxy-acids (a special case of the two functional group variation) were included as test items.

The major difficulty associated with this exercise lay in providing instructions that were unambiguous but that did not contain a self-defeating cue. Perhaps the most precise wording would have been "Write down the number of each cell that contains a compound whose reactions would include the set of characteristic reactions shown by the compound in cell ...". This wording was rejected because it was felt (and the the panel of teachers concurred in this opinion) that it would have been incomprehensible to many pupils. For greater simplicity, reference was made to one specific characteristic reaction (which pupils would have studied). The final wording used was "The compound in cell 13 reacts in a certain way with sodium. Write down the numbers of other compounds that would react with sodium in the same way. The compound in cell 12 reacts with sodium too. Which compounds would react in the same way as compound 12?" Again, teachers were notified of the possible ambiguity in these instructions, and they were encouraged to enlighten pupils who seemed confused by the wording; however, teachers were asked to be very careful not to give specific cues to pupils.

5.132 The Scoring System

Although the primary interest of this investigation was the way in which pupils made their decisions, their overall performance in each exercise was also of interest.

One index of performance that could have been used for each grid exercise was the coefficient of confusion,[77]

$$C = \frac{r}{R} - \frac{w}{W} ;$$

where 'r' represents the number of positive instances chosen by the pupil, 'w' indicates the number of negative instances chosen, and 'R' and 'W' represent total number of positive and negative instances, respectively, in the grid. While a value of $C = \pm 1$ indicates that a pupil has chosen all the positive (negative) instances, and only positive (negative) instances, no other value of C has a unique interpretation, and, in particular, does not indicate whether a pupil has erred by failing to identify positive instances, or by mistakenly choosing negative instances. In the present investigation, the type of mistake made was an important facet of the overall performance, and we felt that this information should be readily obtainable from the scoring index used.

We took the view that a pupil had to make a decision about each of the 15 test items in a grid. He was therefore awarded 1 point for each correct decision he made, giving a maximum score of 15. After the total score, and separated from it by a comma, the number of negative instances incorrectly chosen by a pupil was recorded. When the total number of positive instances in a grid is known, this composite index enables all the important characteristics of a pupil's overall performance to be determined. For example, the first grid (the carboxylic acid grid) contained seven positive instances (including the amino acid). In this case, a score of 10,2 would indicate that 10 correct decisions had been made, 2 negative instances chosen, and that 3 positive instances had not been chosen. From this one could deduce that 4 out of 7 positive instances and 6 out of 8 negative instances had been correctly identified.

Each pupil was awarded a composite index score for each of the first three grids, and two scores for the last exercise - one for each series of compounds chosen.

There were, effectively, only 14 Test Items in the third (alcohol) grid exercise. However, it was decided to retain a maximum score of 15 for this exercise, to facilitate a comparison of results.

For each of the grid exercises, we defined a level of master y, namely, a score of at least 14,0. Given the test materials used, this represented a high level of achievement, but it was felt that pupils achieving this score could be confidently assumed to have used strategies that were based on a level of conceptual understanding equivalent to the specified levels.

5.133 <u>The Card Game</u>

The card game used in phase three required the use of a specially designed Organic Family card deck. Each deck contained seven Organic families - primary, secondary and tertiary alcohols, carboxylic acids, esters, ketones and aldehydes. There were six cards for each family; two cards inscribed with the family name, and four cards having the structural formula of a compound belonging to the family. In addition, there were four 'jokers' - two pairs of cards that bore formulae of two unknown families.

Consideration of the function to be fulfilled by these learning materials suggested that any game to be played with the card deck would have to require pupils to identify cards of the same family, or cards of a named family, at speed. We felt that the requirement of rapid identification would encourage pupils to adopt an efficient gaming strategy - i.e. to choose the functional group as the characteristic property, and to identify it as a unit (and possibly also to note it first, without making a detailed appraisal of the rest of the formula).

Pelmanism (or Memory) and Snap were two games that required pairing at some speed, and offered the additional advantage that their rules would be known by the vast majority of pupils. (This meant that time would not have to be allocated for rule-learning). The pairing rule required pupils to pair two formula cards from the same family, or one formula card with its family name card. Pupils were permitted to challenge an opponent who was thought to have paired cards incorrectly. The set of formula cards could also be used as "flash" cards.

Two decks of cards were prepared. These differed only in that the formula cards supplied for each family were different. Three or four pairs of packs were supplied to both the schools participating in the third stage of the experiment. Teachers were asked to use the two different sets of packs interchangeably, or to select four formula cards at random from the total of eight formula cards per family each time the cards were used. This strategy was designed to minimize the chance that a pupil would become familiar with four particular exemplars of a family; that is, we tried to maintain the situation in which the formulae were essentially "unfamiliar objects". No formula was used in both the Grid Test and a Card deck, and the Card deck formulae contained no functional group variations within a particular family. A small number of formulae had a double bond within the chain, but apart from this, the only variation in environment was the size of chain and the amount of branching.

5.134 Administrative Details

As in the case of the Combined Tests Experiment, the staff members of the Chemistry Departments participating in this investigation very generously offered to administer the tests themselves, although the interviewing was carried out by the author and a colleague.

The test materials - the Grid Tests sheets, and the card decks - were sent to schools in March 1975, so that they could be administered at the conclusion of the Organic Chemistry course.

Because the Organic Chemistry was the last topic to be taught, pupils began revision work immediately this was completed. This enabled the card game to be introduced to the pupils participating in the third stage as a revision exercise. The functions of the card game were discussed with the teachers concerned, and while the suggestion was made that Pelmanism might be the most suitable introductory game (as it allowed slightly more time for decision making), teachers were asked to choose the game(s) that they considered most appropriate for their pupils. The amount of time to be spent on using the cards was not specified precisely; rather, teachers were asked to use the cards during the week between pre- and post-tests for an amount of time that they felt appropriate for revision of one section of Organic Chemistry.

While the investigation was being conducted, one of the schools that had offered to participate in the second phase was unable to provide time for the interviewing of their pupils. The interviews were therefore conducted at only one school, but in spite of the small number of pupils involved, the results of this part of the experiment have been included because they proved to be most illuminating, and suggested possibilities for further study.

Care was taken in organizing the interviews to devise a procedure which would minimize the possibility of interviewer bias influencing pupils' responses. First, the interviewing was carried out by two persons (each interviewer conducting one half of the interviews). Both interviewers followed a written schedule, which gave the exact wording of each question to be asked, and any allowed prompting. Written notes were made of pupils' responses, and any prompting used was recorded.

5.2 The Results

A total of 210 pupils, drawn from 14 schools, participated in the three phases of the experiment. Of these pupils, 164 were involved in Stage (i) only, 12 were interviewed after they had completed the Grid Test, and the remaining 34 took part in Stage (iii) of the investigation.

The experimental results will be presented in four Sections, dealing respectively with pupils' overall performance in the Grid Test, the pattern of their responses in this test, the information obtained during the interview, and finally the results of Stage (iii) of the experiment. The pre-test results of the Stage (iii) participants were included with all other Grid test results in determining pupils' overall performance, and in the analysis of their patterns of response.

5.21 Pupils' Overall Performance

The different aspects of pupils' overall performance in each of the Grid tests are shown in Table 5.1.

TABLE 5.1

PERFORMANCE IN THE FOUR GRID TESTS (N = 210)

	Identity			Behaviour	
	Grid 1 (Acid)	Grid 2 (Ester)	Grid 3 (Alcohol)	Grid 4 (Acid)	(Alcohol)
Mean Score	12.5	10.5	10.9	12.4	11.3
Mean Number -ve Instances Chosen	0.8	1.4	0.8	1.2	0.5
Number Positive Instances in Grid	7	6	7	4	6
Percentage Achieving Mastery	41	6	15	32	6

Very few pupils achieved Mastery (a score of 14,0 or 15) in any but the first Grid test. The mean values shown in Table 5.1 indicate that the low scores were mainly due to a failure to identify correctly positive instances in each grid. In fact, it can be seen that fewer than half the positive instances were identified in Grids 2 and 3, and in the Alcohol series of Grid 4.

It is interesting to compare the percentages achieving mastery in the Grid tests with the percentages of pupils who were successful in identifying family members in the Open Day Experiment (Tables 1.5 and 1.6). As the two experiments differed in format and in the number of positive instances given for each family, the results are not directly comparable. However, it would seem that the present results are superior probably only in the case of the carboxylic acid family, supporting the suggestion made in Chapter 1 that the results of the Open Day experiment were not attributable to the time of year at which the testing was done.

A potentially interesting contribution to the number of negative instances incorrectly chosen was any 'cross-identification' of family members - that is, any identification of an acid, ester or alcohol as a member of one of the other two families. The mean percentages of pupils cross identifying items are shown in Table 5.2.

TABLE 5.2

CROSS-IDENTIFICATION OF FAMILY MEMBERS (AS PERCENTAGES)

Item's Family	Identified as			Reacts as	
	Acid	Ester	Alcohol	Acid	Alcohol
Acid	-	18	6.5	-	8.5
Ester	8	-	N.A.	14.5	4.0
Alcohol	N.A.	N.A.	-	12.5	-

From these figures it is apparent that there were only three cases in which cross-identification was of consequence, namely, acids identified as esters, and esters and alcohols chosen as showing the reactions of an acid.

5.22 Responses to the Grid Tests

Tables 5.3-5.6 show the Test Items and exemplar(s) for each of the grids, and also the percentage of pupils who indicated that each item was the "same" as the Grid exemplars. In the fourth Grid, the first percentage relates to a choice "reacts like compound 13", and the second to a choice "reacts like compound 12", while the third figure in brackets indicates the percentage who decided that the Test Item would show the reactions of both exemplars. So, for example, 59% thought Item 1, Grid IV would show only the acid reaction, 4% only the alcohol

TABLE 5.3

GRID 1: ITEMS AND SELECTION PERCENTAGES

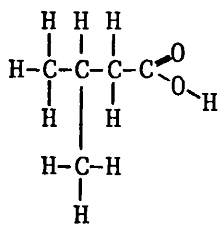
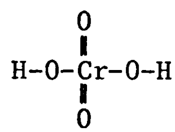
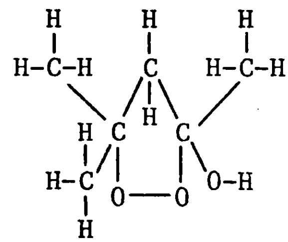
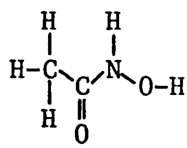
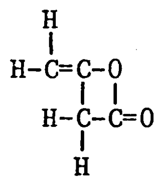
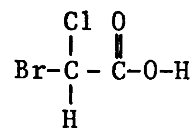
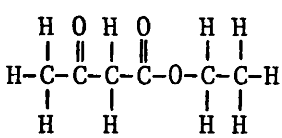
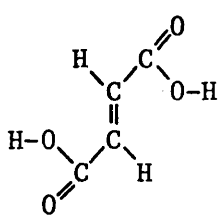
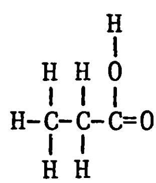
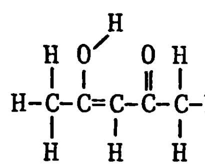
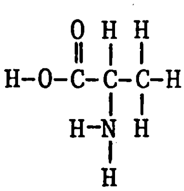
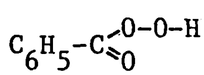
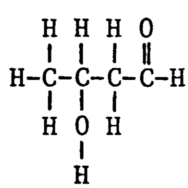
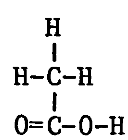
1.  Exemplar	2.  4%	3. $(CH_3)_2CHCO_2H$ *60%	4.  9%
5. $C_5H_{11}COOH$ *79%	6.  7%	7.  5%	8.  *58%
9.  8%	10.  *55%	11.  *87%	12.  13%
13.  *51%	14.  9%	15.  22%	16.  *79%

TABLE 5.4

GRID 2: ITEMS AND SELECTION PERCENTAGES

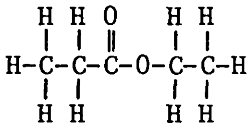
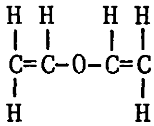
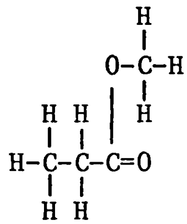
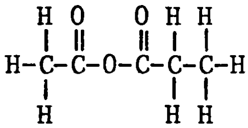
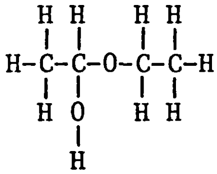
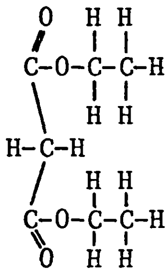
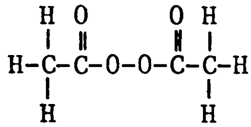
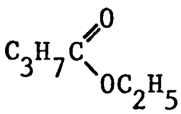
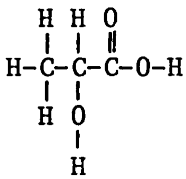
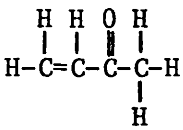
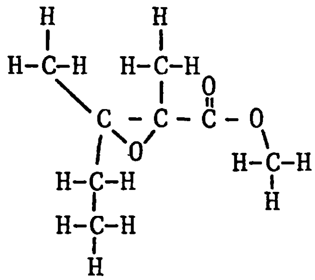
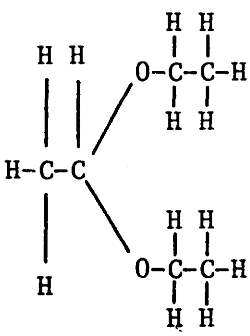
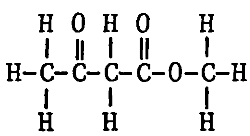
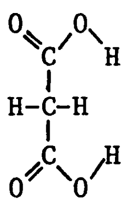
1.  Exemplar	2.  7%	3.  75%*	4. $C_6H_5CO_2CH_3$ 38%*
5.  41%	6.  10%	7.  50%*	8.  28%
9.  70%*	10.  19%	11.  10%	12.  50%*
13.  7%	14.  50%*	15. $C_2H_5OCOCH_3$ 45%*	16.  17%

TABLE 5.5

GRID 3: ITEMS AND SELECTION PERCENTAGES

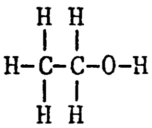
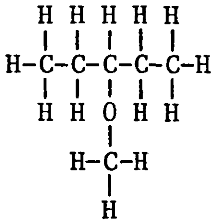
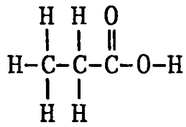
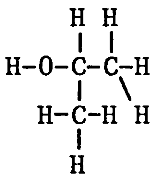
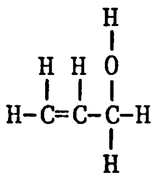
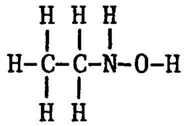
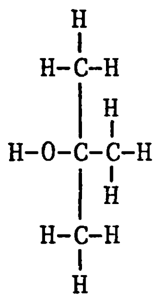
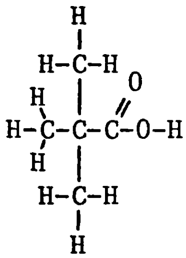
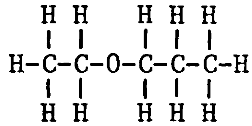
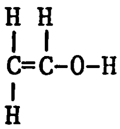
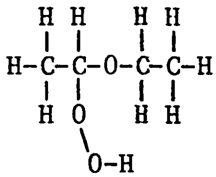
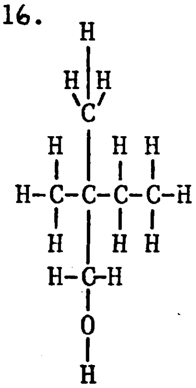
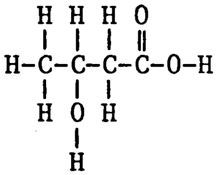
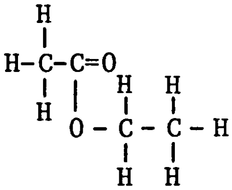
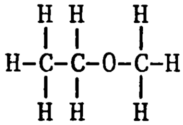
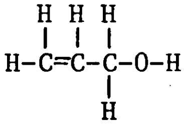
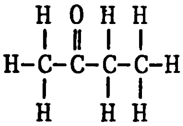
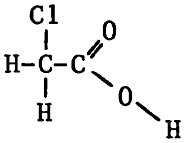
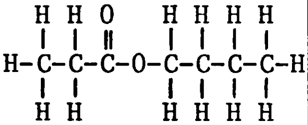
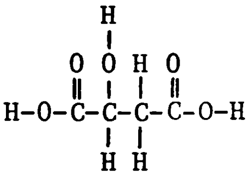
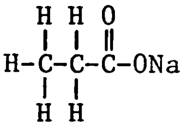
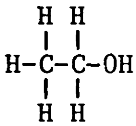
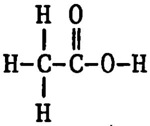
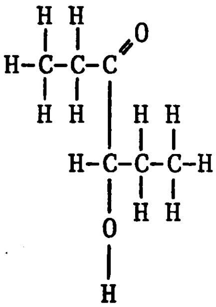
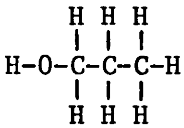
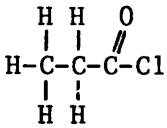
1.  Exemplar	2. CH_2CHOCH_3 *21%	3.  9%	4.  8%
5.  *79%	6.  *50%	7.  15%	8.  *70%
9. Withdrawn	10.  5%	11. $(C_2H_5)_2\ OH^+$ 22%	12.  10%
13.  *51%	14. $CH_3CHOHCH_2CH_3$ *40%	15.  9%	16.  *72%

TABLE 5.6

GRID 4: ITEMS AND SELECTION PERCENTAGES

1.  71%*, 16%*, (12%*)	2.  16%, 5%, (4%)	3.  1%, 10%, (0%)	4.  9%, 50%*, (3%)
5.  7%, 4%, (0%)	6.  72%*, 9%, (5%)	7.  13%, 3%, (0%)	8. $(CH_3)_3COH$ 10%, 65%*, (4%)
9. $C_2H_5CO_2H$ 57%*, 8%, (2%)	10.  44%*, 14%*, (8%*)	11.  11%, 3%, (0%)	12.  4%, - (4%)
13.  -, 4%, (4%)	14.  24%, 29%*, (4%)	15.  7%, 73%*, (3%)	16.  16%, 5%, (1%)

reaction and 12% thought it would show both characteristic reactions. For all the Grids, the percentages choosing positive instances have been indicated by an asterisk.

The variation in percentages selecting each item from the different schools did not seem excessive; nor did there seem to be any between-schools variation in response to different types of Items. This apparent homogeneity of the overall sample justified the pooling of percentages from the different schools. The mean selection percentage and S.D. per School for each Item is shown in Appendix 5.1.

Inspection of the response rates indicated that approximately the same percentages were associated with the different examples of each particular variation. It was therefore appropriate to define a series of effect sizes in terms of the ranges of selection percentages, to provide a scale for ordering the relative effects of the different variations.

Four effect sizes were defined - small, medium, large and very large. The choice of boundary percentages for each range was not completely subjective. For N = 210, $w_{max} \simeq 10\%$. This suggested that no percentage range could sensibly be less than 10%. It also suggested that a range of 0-10% for selection of negative instances could very naturally be equated with a "small effect". Again, the highest percentages associated with positive instances were about 70-80%, which suggested that this range could very reasonably be designated the "small effect" range for positive instances. We took the view that any variation in a positive instance that causeda pupil to <u>guess</u> whether or not an Item should be chosen could well be deemed to have a large effect. Therefore, we equated - for positive instances - the range of percentages not significantly different from 50% with a large effect.

For N = 210, the required range (for $\alpha = .05$) is 40-60%.

Using these guidelines as a basis the effect sizes were equated with selection ranges as follows:

Effect Size	Selection of Positive Instances	Selection of Negative Instances
Small	70-80%	0-9%
Medium	60-69%	10-19%
Large	40-59%	20-29%
Very large	Below 40%	Above 30%

The variations of functional group and of environment associated with each effect size are shown in Table 5.7 and 5.8 respectively. From these results it would seem that many pupils considered functional groups to be collections of bits, rather than units, to a significant extent.

The only significant misidentifications were ester for acid and vice-versa, and ether for alcohol - that is, misidentifications were made between $-C{\overset{O}{\diagup}}$ -O-H and $-C{\overset{O}{/\!\!/}}$ -O-R, and between -O-H and -O-R. Such mistakes could arise if the functional groups were matched bit by bit, and 'R' and 'H' were mistakenly equated.

The fact that separation of group elements gave rise to a large effect is also significant in this respect. (Item 12, Grid 1 drew a smaller response than the other examples of this variation. However, this Item contained a double bond in the chain, a variation that gave a large effect on its own, and which could well have depressed the selection percentage for this Item).

TABLE 5.7

EFFECT OF VARIATIONS IN FUNCTIONAL GROUP

Classification by Effect Size	Items
Small Effect	
(a) Orientation of the Functional Group.	I:11,16. II:3,9. III:5,8,16. IV: 15.
(b) "Trick" Functional Group: additional element.	I:14. III:15.
(c) One group identified as another, EXCEPT acid as ester and ester as acid (Medium Effect).	
Small to Medium Effect	
(a) Non-structural representation: -COOH, -OH at the end of a chain.	I:5. IV:8
(b) Ether identified as alcohol.	III:3,12. IV:3.
Large Effect	
(a) Two identical Functional Groups.	I:10. II:7. IV:10.
(b) "Trick" Functional Group: separation of Group elements.	I:12,15. IV:14.
(c) Non-structural representation: $-CO_2H$.	I:3. IV:9.
Very Large Effect	
(a) Non-structural representation: $-CO_2-$, -OCO-, -OH-, -HO-, within a chain.	II:4,15. III:2,14.
(b) "Trick" Functional Group: replication of all or part of a Group.	II:5,8.

Note: Roman Numerals denote Grid numbers

TABLE 5.8

EFFECT OF VARIATIONS IN ENVIRONMENT

Classification by Effect Size	Items
Small Effect	
(a) Replacement of one C.	I:2,6. III:7.
(b) Replacement of one H.	IV:6.
(c) Size and amount of branching of chain.	I:11,16. II:3. III:5,8,16. IV:15.
Medium Effect	
(a) Replacement of two H's.	I:8.
Large Effect	
(a) Inclusion of a double bond in C-H chain.	III:6,13. IV:4.
Large to Very Large Effect	
(a) Inclusion of another Functional Group. (* for alcohol series)	I:13. II:12,14. IV:1*,10*,14*.

Note: Roman Numerals denote Grid numbers

While inclusion of an additional element in a functional group, and replacement of an adjacent C had only a small effect, replication of all or part of the (ester) functional group had a very large effect. In fact 17% of the sample selected all of Items 5, 7 and 8 (Grid 2) as esters; interestingly, this was the same as the percentage who selected 7 (in which the group was replicated) but not Items 5 and 8. The increasing effect of these variations also suggested that bit by bit matching of the functional groups was a widely used strategy.

The second point to be considered in the light of these results was whether pupils used the functional groups as characteristic properties - or alternatively, whether students who treated functional groups as collections of bits, considered these bits particularly significant in determining family membership and characteristic behaviour. Although no clear cut conclusion could be drawn concerning this point, the balance of evidence favoured the proposition that pupils did not regard the functional group as the characteristic property (or its bits as particularly significant). The fact that there were generally few cross-identifications of family members was the major evidence supporting the view that pupils did consider functional groups to be of particular importance. There were, however, several results that suggested that the contrary behaviour was prevalent.

The fact that orientation of the functional group, and the size and amount of branching of the chain showed only a small effect suggested that these differences between exemplars and Test Items did not distract pupils significantly. However, almost all other differences between exemplar and Test Item seemed equally significant in affecting pupils' responses.

The percentage of pupils selecting Items having two identical functional groups fell within the random guessing range. It is not unreasonable that fifth year pupils should guess the family of such Items. However, Items with two different functional groups (one exemplar plus one other) drew very similar response rates. We would have expected different response rates for these two types of Item (specifically, we felt that pupils would have considered the former Items more likely to belong to the exemplar family than the latter) if pupils considered the functional group to be the characteristic property.

The effect of a double bond within the chain was of the same order of magnitude as these two variations, again contrary to what one would have expected if the functional group were treated as the characteristic property. (The effect of the double bond may have been due to a carry-over from alkene-alkane learning).

It can also be seen that the effect of replacing 2 hydrogens was much greater than the effect of replacing just one, suggesting that the amount of variation, not just the type of variation, was a significant determiner of pupils' responses.

The responses given to Items 1 and 10, Grid 4 - the hydroxy-acids - were consistent with pupils equating behaviour with a molecule or family, rather than a functional group, (and hence with a failure to consider the functional group as a characteristic property). Very few pupils selected these items as showing the characteristic reaction of the alcohol exemplar. It would be very difficult to argue that this was due to apparent ambiguity in the instructions for this Grid; one would have expected such ambiguity to produce very low response rates for both the 'like acid' and 'like alcohol' series. A perfectly feasible explanation of this result is that it arose because the Items that showed the characteristic carboxylic acid reaction had to be listed first, and so pupils may have decided that as Molecules 1 and 10 behaved 'like acids' they could not also behave 'like alcohols'. (This line of reasoning could also have contributed to the very low response rate for Item 14, Grid 4, which contained a 'trick acid' group). It should be noted however, that these results do not support this explanation exclusively. If this experiment were repeated, it would be most interesting to ask half the sample to select the Items that reacted like the alcohol exemplar first. If this procedure resulted in the inversion of response rates for the acid and alcohol series hydroxy-acids, strong evidence for the equation of behaviour with a molecule rather than a functional group would have been obtained. However, it could be that the observed responses are due to some arbitrary or idiosyncratic misconception; certainly, this cannot be ruled out on the basis of the present data.

Overall then, the results suggested that the functional group was considered as a collection of bits to a significant extent, and they were certainly consistent with a significant failure to treat the functional group as the characteristic property. It seemed that pupils very probably used a strategy of matching molecules bit by bit, with no great distinction between environment bits and functional group bits, and that the only differences they considered trivial were the size of the chain and its degree of branching, and the orientation of the functional group. In fact, considering that the effect of the non-structural representation COOH was less than CO_2H, that -OCO- was less than $-CO_2-$, and that in all cases a non-structural representation of a functional group within a chain had a greater effect than a similar representation at the end of a chain, the matching technique may have been very literal indeed.

These conclusions were supported by the results of the interview, which are reported in the next section.

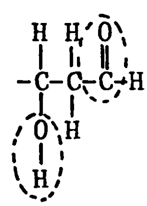
5.23 The Results of the Interview

Summaries of pupils' responses to the interview questions have been included in Appendix 5.2.

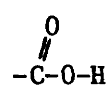
In answering the three questions 'If someone asked you what a .. (family) .. was, what would you say?', the majority of pupils defined carboxylic acids and alcohols in terms of the functional group, but only four attempted to define an ester in this way (and one of these referred to a 'COCH' group). Half the pupils described an ester as 'a compound between an acid and an alcohol'. This suggested that pupils might well have been repeating a given definition of the acid and alcohol families, and certainly suggested that they did not always consider the functional group to be a defining or characteristic property of a family. Interestingly, two candidates (10 and 5) referred specifically to "bits" of the COOH group.

In comparing NaOH, H_2O and CH_3OH half the pupils commented that all contained an -OH group. No student felt that this should indicate a behavioural relationship between the compounds, even if specifically questioned on this point. In fact, the reaction was "They're different compounds, so they are different".

In discussing their responses to the Grid sheets, some pupils commented on practically all Items on their chosen sheet, but many seemed able - or prepared - to give an explicit reason for rejection of only a few Items. Two or three pupils said that they were looking for a certain group or family, but only one of these (No. 5) actually seemed to use a functional group criterion. Generally, the other pupils did not seem to consider the functional group and the environment separately, and then reach a decision based on the functional group, but regarded the molecule as a whole and tried to match it piece by piece. Several pupils spoke of 'cancelling' the different parts, and many stated that you could cancel any number of CH_2 units for any number in the exemplar. This matching was applied equally to the functional groups - giving rise to comments like "it's missing $\overset{\text{O}}{/\!/}$" (an alcohol compared to an acid), or "the OH isn't matched" (a ketone compared to an acid), where one might have expected 'it's got the wrong group'. Several pupils were puzzled by double bonds in the C-H chain, specifically saying that they didn't know whether one could cancel these for 'ordinary' bonds. One pupil (candidate 6) who had recorded Item 15, Grid 1 as a positive instance commented that he had become unsure of that choice as he was not sure that one could cancel



for

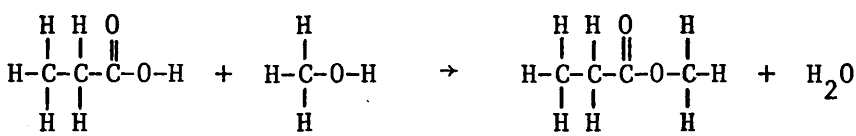


Some arbitrary criteria were used (e.g. "alcohols don't have double bonds", "acids don't have rings"); these were also consistent with the use of bit by bit matching, and gave no indication that functional groups were being used as the characteristic property.

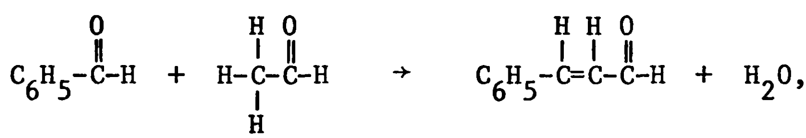
The last two Interview questions which investigated pupils' interpretation of "same" and "different" in the context of esterification and condensation reactions, and the methods and criteria used in reaching their decisions, proved to be very difficult for most pupils, who took a long time answering them. (This was not unexpected; during the original classroom observation many pupils had said that big molecules and big equations were hard). The second reaction pair caused considerably more difficulty than the first and most pupils said that the equations and formulae were hard.

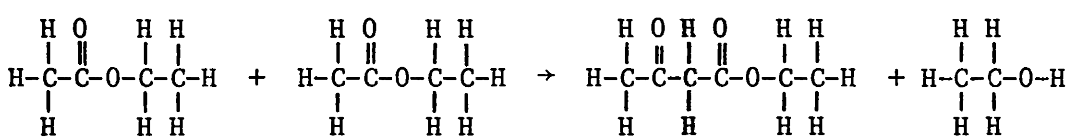
First, pupils were shown the pair of equations:

$$HCl + NaOH \rightarrow NaCl + H_2O,$$



and then the equations:





In answering these questions, most pupils seemed to set up an implicit correspondence between pairs of reactants and products in the order written, thus:

$$\begin{array}{ccccccc} A & + & B & \rightarrow & C & + & D \\ \updownarrow & & \updownarrow & & \updownarrow & & \updownarrow \\ A' & + & B' & \rightarrow & C' & + & D' \end{array}$$

Many pupils began by attempting to name each of the compounds, or their families (a slow and laborious task). Some then named the "family pattern" - e.g. "Acid + Alkali gives Salt + Water". Some students compared the two patterns - suggesting an equation of behaviour with family, rather than functional group. The strategy generally used was to compare A to A', B to B', C to C' and D to D'. The criterion used by most pupils was evidently that a pair of reactions would be "the same" to the extent that the reactant and product pairs were "the same" - again suggesting that behaviour was not seen as a direct correlate of the functional group.

In a number of cases a comparison was made of the families to which a corresponding pair of compounds belonged. Where a comparison of the formulae was made, a general bit by bit matching was again evident. Only Candidates 5 and 6 seemed to go beyond a comparison of the reactants and products.

With the possible exception of these two candidates, it seemed that at best, pupils would achieve success only if they were presented with two cases having the same family reaction pattern. In effect, they did not compare the reactions at all, and their method of comparing compound to compound would make it very unlikely that they would at any stage deduce the nature of any reaction involving unfamiliar compounds not susceptible to family patterning, (and many condensation reactions would fall within this latter class), let alone be able to compare two such reactions.

The mean scores obtained in the Grid test by the pupils who were interviewed are shown in Table 5.9.

TABLE 5.9

MEAN GRID TEST SCORES FOR THE INTERVIEW SAMPLE

Grid 1	Grid 2	Grid 3	Grid 4	
(Acid)	(Ester)	(Alcohol)	(Acid)	(Alcohol)
12.0, 1.3	10.5, 1.7	9.9, 0.5	12.3, 1.1	11.9, 0.3

A comparison of Tables 5.1 and 5.9 suggests that this group of pupils were average in respect of their overall performance. In spite of this, it would not be wise to generalize too widely from the results of the interview, due to the small sample size involved. However, the results certainly tended to support the interpretation of the Grid Test results, and gave weight to the propositions that pupils often did not treat functional groups as units, that they were certainly not greatly committed to the use of a functional group as a characteristic property, and that they tended not to relate behaviour specifically to the functional group.

5.24 Stage (iii): The Pre- and Post-Test Results

Of the 34 pupils who completed the pre-test, 28 returned post-tests. The pre- and post-test mean overall scores for these pupils are shown in Table 5.10. (The figures in brackets refer to the choice of negative instances).

TABLE 5.10

PRE- AND POST-TEST RESULTS (N = 28)

Grid	Pre-Test Means	Post-Test Means	Mean Difference	95% Confidence Interval
Acid	12.65 (.76)	14.25 (.11)	1.64 (.65)	{0.8, 2.5} {0.13, 1.17}
Ester	10.35 (1.64)	12.28 (.82)	1.93 (.82)	{1.04, 2.82} {0.34, 1.3}
Alcohol	11.46 (.44)	12.74 (.63)	1.3 (-.19)	{0.7, 1.9} {-0.53, 0.15}
Acid Behaviour	12.72 (.72)	13.82 (.32)	1.11 (.4)	{0.73, 1.5} {0.09, 0.71}
Alcohol Behaviour	11.35 (.27)	12.65 (.12)	1.31 (.15)	{0.52, 2.1} {-0.15, 0.45}

These figures indicate that an improvement in overall performance occurred in each case, and that this was significant statistically and practically. In most cases, a significant decrease in the mean number of negative instances chosen also occurred, and, on average, pupils identified at least 2/3 of the positive instances in a grid correctly in the post-test.

More detailed information about the improvement in performance may be obtained by consideration of the pre- and post-test selection percentages for each Test Item, shown in Table 5.11. (The percentages for Items in Grid 4 again follow the order "like acid", "like alcohol", with the percentage for "like both" shown in brackets).

The improvement in selection of positive instances and rejection of negative instances indicated by the overall performance scores was also clearly evident in the differences between pre- and post-test selection percentages for individual Items. (As a guide for these comparisons, $w_{max} \simeq 26\%$ ($\alpha = .05$) and 16% ($\alpha = .1$) for N = 28). An overall impression of the relative effects of the different variations in functional group and environment in the pre- and post-tests may be obtained by noting that for positive instances, selection percentages fell outwith the small effect size for 7 Items in the post-test compared to 21 Items in the pre-test, while for negative instances there were 4 such results in the post-test compared to 12 in the pre-test.

There was a decrease in the cross-identifications of both acids and esters, and of ethers as alcohols, and also a fairly considerable decrease (from medium to small) in the effect of separation of group elements. (See Tables 5.7 and 5.8 for a listing of variation examples). Generally, the effect of a non-structural representation of a functional

TABLE 5.11

PRE- AND POST-TEST SELECTION PERCENTAGES (N = 28)

	Grid 1				Grid 2				Grid 3			
Pre-test		10	54*	7		7	68*	39*		7*	4	7
	79*	36	0	64*	46	18	50*	36	89*	54*	4	78*
	7	68*	89*	18	82*	14	14	50*	-	4	15	7
	43*	10	18	86*	4	64*	46*	25	41*	33*	4	82*
Post-test		4	82*	0		4	97*	57*		18*	4	0
	97*	0	0	75*	50	4	71*	18	97*	85*	18	97*
	0	93*	97*	0	97*	0	7	71*	-	0	22	7
	86*	4	4	97*	4	82*	46*	10	75*	64*	0	100*

Grid 4

Pre-Test								Post-Test							
64*	14*	11	7	4	7	4	54*	71*	38*	7	0	0	4	0	96*
(12)*						(4)		(35)*							
0	4	82*	0	7	4	7	71*	4	0	86*	0	7	0	0	88*
						(4)									
54*	0	43*	7*	4	4	7	/	86*	0	64*	23*	0	0	0	/
		(8)*				(7)				(19)*					
/	4	14	25*	7	93*	7	4	/	4	7	50*	0	92*	7	4
(4)				(4)				(4)							

group decreased considerably, although the within-chain instances for the ester and alcohol groups were still not within the small effect range in the post-test.

These changes suggested that pupils were showing a greater tendency to regard the functional group as a unit. However, it should be noted that replication of all or part of the (ester) functional group showed a medium to large effect in the post-test with improvement shown for only one of the two examples of this variation.

There was a clearer suggestion that pupils had moved towards the choice of the functional group as a characteristic property in the post-test. The effects of a double bond in the chain, of two identical functional groups and of two different functional groups (in Grids 1-3), and the replacement of two hydrogens, decreased markedly to show only a small effect in the post-test.

While there was a considerable increase in the percentages of pupils selecting the hydroxy-acids as showing the reactions characteristic of an alcohol, these response rates were still low in the post-test. This suggested that there was a decreased, but still significant, failure to relate behaviour specifically to functional groups.

Overall then, the pre- and post-test results showed that pupils' performance had improved after using the learning materials, and that they had certainly shown a greater tendency to use the functional group as the characteristic property, and an increased tendency to treat a functional group as a unit. It should perhaps be emphasized that the card decks did not contain the sort of variations used in the Test Items. Therefore, the changes in pupils' behaviour could not be seen as a result of increased practice with formulae very similar to the Test Items, but could be attributed to changes in strategies.

5.25 Conclusions

Because of the generalisability of the Combined Test results, it could be assumed that the majority of pupils who participated in the Grid Test would not have recognized as units the Class I groups (Section 3.44). In addition, the results of the first two stages of the Grid Test Experiment indicated that:

(i) There was a significant failure to treat the functional group as a unit;

(ii) There was very probably a failure to choose the functional group as a characteristic property;

(iii) There was a clear tendency to match formulae bit by bit, without necessarily making any great distinction between functional group and environment;

(iv) There was very probably a failure to relate behaviour specifically to functional groups.

Taken together, these results provided strong evidence in favour of there being a lack of the specified conceptual understanding of functional groups. As we have seen, the mistakes pupils made in the (typical) tasks they were required to perform in the Grid Test Experiment followed a consistent pattern, and so were directly relatable to a lack of conceptual understanding.

The third stage of the Experiment indicated that the effect of this lack of conceptual understanding was certainly reduced by the use of the specially designed learning materials.

Thus, the results of the Combined Tests and the Grid Test Experiments gave strong presumptive evidence for the validity of the hypothesis "That pupils' difficulties were due to a lack of conceptual understanding of functional groups, (at the specified levels or states)".

This conclusion posed a very interesting question. Given that conceptual understanding of functional groups could well be expected to be very relevant in almost all areas of Organic Chemistry, why had pupils selected C.H.E. reactions in particular as an area of difficulty? We will propose a possible answer to this question in the next Section.

5.3 An Hypothesis Relating Information Content, Conceptual Understanding and Difficulty - The I.C.C.U.D. Hypothesis

The comment which had been made again and again at different stages of the investigations reported was that "big" molecules and "big" equations were hard. It seemed to be the size or information content of ester and condensation formulae and equations which differentiated this topic from other areas of Organic Chemistry. The question thus arose, given a low level of conceptual understanding, would the adjudged difficulty of a topic depend on the (magnitude of the) information content?

We have already seen (Section 3.3) that the limited storage capacity of Immediate Memory constrains the amount of information that can be stored and retrieved in an immediate recall task (such as the reproduction of patterns and formulae in the Combined Tests Experiment). This storage of information is an <u>active</u>, not passive procedure, requiring processes such as perception, recognition, chunking and coding, for example, to act upon the given information.

Immediate Memory has been called "Working Memory",[78] a term which emphasizes rather nicely the (storing + processing) nature of the system.

It would seem that, in addition to its limited storage capacity, there is a limitation on the processing capacity of Working Memory.

For one thing, the processing is performed in a time-sharing mode rather than simultaneously; if one process is particularly time consuming, it can prevent or detrimentally affect, the performance of other processes. Massaro[79] cites an example that is commonly observed; when young children read a text aloud, they may be quite unable to give the meaning of a sentence because most of their available processing time is taken up with perceiving the words and phrases to be spoken. When processes aimed at eliciting the meaning of the text could begin, earlier essential words have passed from Working Memory.

While in the case of a simple immediate recall task, it is useful and appropriate to describe the nett constraint on Working Memory as a storage capacity (of approximately 7 $\pm$ 2 chunks of information), in considering more complex tasks (such as the deduction of a reaction mechanism from a given equation) it is very difficult to specify the capacity of Working Memory as X chunks of information storage + Y amount of processing. For one thing, it can be very difficult to draw a clear distinction between the "storage" and "processing" components of a task. For example, chunking may reduce the storage requirement of a task, but will, at the same time, increase its processing load.

Following a series of very interesting experiments, in which students were required simultaneously to store information for subsequent recall and to perform cognitive activities such as comprehension and reasoning, Baddeley and Hitch[80] have proposed a model of Working Memory that allows a partial trade-off between storage load and processing capability. They have suggested that the Working Memory system may contain a component which is used only for storage, and a "flexible work space" or "central processing space" which can either supplement the storage capacity of the "store only" component, or can be used for processing. Their experiments indicated that the capacity of the "store only" component was the traditional Immediate Memory capacity (i.e. 7 $\pm$ 2 chunks) and that it was only when the required memory load exceeded this capacity that they observed performance decrement (in terms of the recall of stored information and the success of the comprehension or reasoning tasks). In such a situation, according to their model, the central processing space must accommodate the information overflow, and the decreased space available must cope with the additional processing load of "servicing" and retrieving the stored information, as well as maintaining the cognitive processing required by the imposed task.

While we do not intend to make any detailed use of this particular model of Working Memory, we would suggest that when a pupil is given some information (an equation, or the text of a problem, for example), and asked to deduce or infer from it some new information, the limited capacity of Working Memory constrains the amount of information he can handle simultaneously in the performance of this task. In contrast to an immediate recall task, the pupil will not necessarily have to store all the given information; on the other hand, he may have to hold in working memory additional information - such as relational information - deduced from the given information, or retrieved from Long Term Memory. The essential question is whether, given the constraints of Working Memory he can store and operate on sufficient information to be successful in the required task.

If we consider the Immediate Memory span to be the capacity of Working Memory under the condition of a minimal processing load, it is certainly reasonable to expect that performance will be impaired, and success jeopardized, if at any time a pupil attempts, or is required, to exceed this storage capacity, whilst also operating on the stored information. We would also suggest that the level of (relevant) conceptual understanding is an important variable in determining whether or not a pupil will be successful, in that it affects both his capacity to store information and his ability to operate successfully on this information.

In more detail, we would suggest that information content of tasks, related conceptual understanding and adjudged difficulty are related in the following way:

1. The number of chunks represented by a given body of information will depend upon the level of relevant conceptual understanding.
2. The larger the number of chunks (from given information, together with any additional chunks) required at some stage in the task, the greater its adjudged difficulty, and the poorer the observed performance.
3. In the limit, if chunk capacity is exceeded, no useful information may be extracted if the pupil attempts to handle the given information 'in-a-one'.
4. When chunk capacity is exceeded, new information may be obtained by using a memory conserving strategy which allows a sequential consideration of the information.
5. Conceptual understanding leads to an efficient (small number of steps) organized (steps performed in an efficient order) converging (leading to a synthesis of information) strategy. Lack of conceptual understanding can lead to an inefficient (large number of steps) poorly organized strategy, or even an arbitrary or diverging strategy.

This proposed relation between Information Content, Conceptual Understanding and Difficulty will be referred to as the I.C.C.U.D. Hypothesis. In practical terms, the hypothesis proposes that where there is a lack of conceptual understanding, pupils may perform reasonably (while not necessarily showing mastery) and not complain of difficulty in low information situations, but in high information situations - i.e. situations in which, at some stage, the expected number of chunks exceeds chunk capacity - performance will drop dramatically, and pupils will complain of difficulty. It is a familiar complaint of teachers that their pupils can handle the isolated bits of a problem, but can't seem to put them all together.

Much of the work on Learning, Chunking and Memory reported in the literature is not directly relevant to the propositions of the I.C.C.U.D. Hypothesis, in that an emphasis is placed upon recall of presented material, rather than extraction of new material. One very interesting exception is a study performed by Wanner and Shiner.[81] Their starting position was that, in solving a mental arithmetic problem, students will often have to store some information temporarily for use at a later stage of the solution. They refer to such information loads as "Transient Memory loads", and argue that "the solution to a problem may break down if its transient memory load exceeds the limited capacity of Short Term Memory".

In a series of experiments, they required students to solve 'left parenthesis' (L.P.) and 'right parenthesis' (R.P.) mental arithmetic problems, such as those shown in Figure 5.1. For both problem types, subjects were shown, sequentially, the five segments of information into which problems were divided. At one of the three break points (indicated by asterisks in Fig. 5.1) the display sequence was interrupted by a list of five names that subjects were required to memorize. Wanner and Shiner argued that the transient memory loads imposed by the arithmetic task at break points 1 and 3 were similar for R.P. and L.P. problems, but that the transient memory load at break point 2 was greater for R.P. than L.P. problems. They therefore predicted that performance on the arithmetic task and on the recall task would be similar for both problem types, when the list of names was interpolated at points 1 or 3, but that R.P. results would be significantly worse than L.P. results when the interpolation occurred at point 2. This prediction was confirmed by their results. They suggested that at the second break point, the transient memory load for the R.P. problem illustrated in Fig.5.1 would be "5 - (4"; this, together with a list of 5 names could well be expected to exceed chunk capacity. For L.P. problems, however, transient memory load at break point 2 would be only "1" so that under the L.P. condition chunk capacity would not necessarily have been exceeded. This result of Wanner and Shiner would thus illustrate proposition 3 above. Interestingly, Wanner and Shiner reported that only 3% of their subjects indicated that they employed the strategy "change all signs after the -" in solving R.P. problems (i.e. converting the illustrated R.P. problem to the L.P. problem 5 - 4 + 1). Such a strategy would certainly decrease the transient memory load at break point 2; we would argue that such a strategy would be used by someone with a good conceptual understanding of subtraction.

Segments of Information	(5	-	4)	-	1	(L.P.)
Break points	*1		*2		*3	
Segments of information	5	-	(4	-	1)	(R.P.)
Break points	*1		*2		*3	

Figure 5.1 L.P. and R.P. versions of an arithmetic problem

In a rather different area, Yngve[82] has argued cogently, (and with some very entertaining examples) that the rules of English grammar and of good usage operate to minimize the amount of information that must be retained in Working Memory at any one time in order to extract meaning from the written text. In other words, these rules seem to take implicit account of the limited capacity of Working Memory.

To clarify the relationships proposed by the I.C.C.U.D. hypothesis, the points listed above will be illustrated in terms of Organic Chemistry and the observed behaviour and performance of pupils.

5.31 <u>The Number of Chunks Representing a Given Body of Information</u>

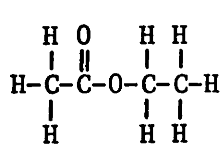
Conceptual understanding can affect the number of chunks in three ways:

(i) By increasing the amount of information per chunk;

(ii) By implicitly declaring some of the information redundant in the first instance at least. That is to say, salient information is readily identified, and stored; the remaining, 'redundant', information is disregarded, because it has been deemed unlikely to contribute to the first phase of solution at least;

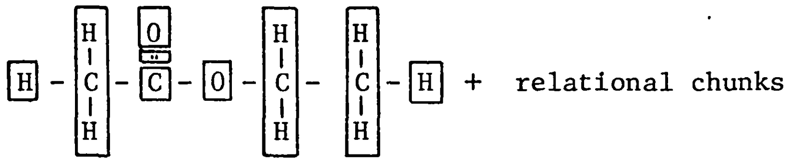
(iii) By allowing the combination of some of the given information with some additional information to form a chunk.

Consider one of the formulae used in the second pair of equations in the interview,



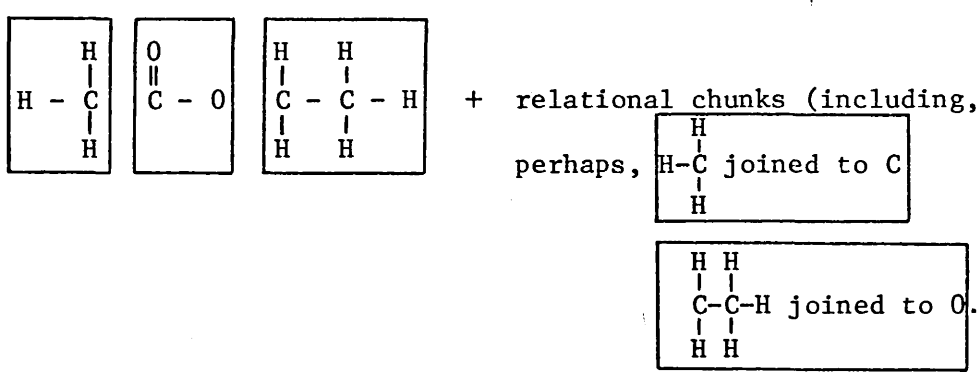
.

(i) Given very low conceptual understanding of functional groups, this could be chunked as:

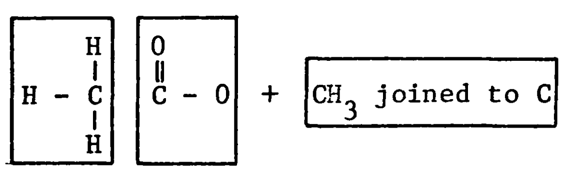


(This corresponds roughly to the recognition of Class III groups as units, but does not indicate the memory saving that would accrue if account were taken of replication of units).

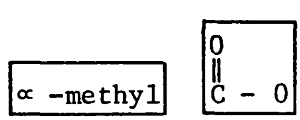
(ii) Better conceptual understanding of functional groups could increase the information content per chunk:



(iii) The C_2H_5 could be declared redundant initially:



(iv) Incorporating a relation with some given information could give:



Stage (ii), which could be possible at the secondary level, represents a considerable reduction in the number of chunks required in comparison with Stage (i). Although pupils at this level would not be expected to carry out a Stage (iii) reduction, they might realize that the precise details of a multi-branched chain were not essential, and code or chunk it as 'big chain', for example. The implicit declaration of some information as redundant is, we would suggest, an important consequence of conceptual understanding, and may reduce the memory load significantly.

Where pupils complained of big molecules' and formulae's being hard, it was quite clear that they referred to the totality of the information, rather than any chemical complexity. Some pupils seemed to feel that they had to store the chain of a formula precisely, which was almost impossible, given their inefficient chunking. Perhaps for this reason, many pupils exhibited a strong desire to name a compound represented by a formula (a coding device), even when this was unnecessary and unproductive. It was noticeable that the interview candidates (5 and 6) who gave good answers for the second reaction comparison ignored the precise details of the chains.

5.32 The Use of Strategies

In the Combined Tests Experiment, the mean fifth year Molecule Score was about 10 - i.e. the 'average' pupil could reproduce correctly formulae up to Difficulty 10 (Section 4). Thus, a Difficulty 10 formula gives an estimate of the average Short Term Memory capacity for this type of information given the chunking corresponding to "average" conceptual understanding of functional groups. A typical Difficulty 10 formula is shown in Figure 5.2.

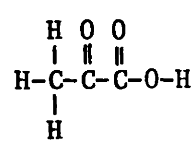


Figure 5.2

Comparison of this formula with those in the Grid Test (Tables 5.3-5.6) suggests that many of the latter would be at or beyond the capacity of Short Term Memory. However, reasonably simple strategies were available for classifying by family. Pupils with a reasonable level of conceptual understanding could examine the functional group of a Test Item, then either reject the Item, or check the environment before accepting it. The results of the Test suggested that some pupils had used a strategy based on low conceptual understanding, in which they compared bits of the functional group and environment. The latter strategy would require more steps than the former, but not an excessive number. Furthermore, there would be no need to synthesize the outputs of each step - the comparisons would continue until one gave the output "different", at which stage a decision could be reached on the basis of that output alone.

Thus, although inefficient strategies (due to low conceptual understanding) should lead to less than mastery performance, it would not be expected that the information content of this situation would lead to a designation of "difficult" for this task. Pupils in the interview sometimes appeared uncertain of the correctness of their decisions, but they did not regard the task itself as difficult; this was not suggested by pupils in schools either.

Where pupils have to interpret a C.H.E. equation, or compare two reactions, as in the interview, the situation is different. Again, each formula in an equation may well exceed chunk capacity. A pupil with good conceptual understanding could, in some cases, handle all the essential information "in-a-one" and extract the required new information. If this were not possible, he would be in a position to use an efficient, organised, <u>converging</u> strategy - perhaps considering the functional groups first and then their environments if necessary, synthesizing the successive outputs. A pupil with low conceptual understanding would almost certainly not be able to handle the information he considered necessary "in-a-one", and so he would be completely dependent on the use of a memory conserving strategy. However, by virtue of his low conceptual understanding, he would be much less likely to be able to adopt a strategy with a high probability of success.

The strategy observed in the interview - naming each compound, deducing its family, and then naming the family reaction pattern, may be adequate (if inefficient) in some circumstances, but not in all. For example, where pupils were asked to compare and contrast two reactions, this strategy was extended to make the A-A' etc. comparison described above. This was clearly a diverging strategy, as it led pupils away from a consideration of either reaction, let alone a comparison of the two. (The use of these two strategies suggests a low level of conceptual understanding of chemical reaction as well as of functional groups - quantities that are clearly not independent). A pupil with good conceptual understanding of functional groups would be able to adopt a successful strategy - deducing the reaction shown by each equation, and then making the necessary comparison.

As a second example, consider a single condensation or polymerisation equation. No 'family reaction pattern' can be specified for such reactions, and so a pupil will be successful in extracting information only if he adopts a strategy of observing the change in reactants, rather than the reactants themselves - i.e. the type of strategy associated with good conceptual understanding of functional groups.

Thus, the I.C.C.U.D. hypothesis gives a model which can explain why pupils specified the topic of C.H.E. reactions as difficult. It also points to a vicious circle in the learning of Organic Chemistry. Given that a pupil is introduced to a concept at a suitable time in his stage of development, he may form a low level of conceptual understanding reasonably quickly. However, conceptual understanding is acquired only over a period of time, during which a pupil is exposed to many instances, examples and situations in which the concept is important. In Organic Chemistry, for instance, conceptual understanding of functional groups will be built up by examining family behaviour, reactions, etc. (The Combined Tests Experiment suggested that, for many pupils, the duration of secondary education was insufficient for the emergence of this understanding, given the extant learning conditions).

It is, however, this very conceptual understanding which is so necessary for the efficient and successful extraction of new information, from information given to provide the understanding. Part of conceptual understanding, we have already argued, is an awareness of the importance of the concept; that is appreciation of why a certain quantity, such as a functional group, has been defined and considered so essential. Very often, this importance becomes apparent only in high information situations (such as the condensation reaction considered above), and it is in this type of situation that the conceptual understanding is most necessary as a prior requisite.

Because the three quantities, information content, conceptual understanding and difficulty are perfectly general, the I.C.C.U.D. hypothesis would predict that in other areas of Chemistry one should find the same pattern of some competence in low information situations, combined with very poor performance and a judgement of difficulty in high information contexts, where a low level of conceptual understanding exists.

In the next Chapter, some results from three independent investigations into other "areas of difficulty" in Chemistry will be presented and discussed. These results will be examined for evidence of patterns of behaviour consistent with the I.C.C.U.D. hypothesis.

APPENDIX 5.1

School Selection Percentages: Mean and S.D.

Grid 1				Grid 2				Grid 3			
71.4	3.2	59.7	8.8	32.1	6.7	74.8	37.7	92.1	21.4	9.2	7.6
28.5	7.5	24.0	17.0	24.1	9.5	16.0	18.9	12.9	14.3	10.3	7.0
77.2	6.7	5.3	57.4	41.5	10.4	47.8	28.4	78.8	49.1	15.6	69.8
20.8	8.4	8.0	24.4	16.3	11.3	22.3	16.5	19.1	27.9	12.2	22.9
8.5	54.2	85.9	13.0	70.0	19.9	10.5	50.0	W.D.	4.5	22.0	10.1
13.2	21.0	17.8	10.9	20.6	12.2	6.3	21.7		6.6	13.6	11.4
50.6	9.0	23.0	78.9	7.1	49.4	44.5	16.8	51.4	39.4	8.8	71.6
21.4	11.9	19.0	17.8	6.5	19.0	15.8	13.7	23.5	19.5	7.4	19.4

Grid 4							
Acid Series				Alcohol Series			
71.0	16.0	1.4	9.1	15.4	4.5	10.8	52.4
15.5	14.6	3.9	7.9	11.4	6.4	8.2	21.7
7.6	72.5	13.6	10.4	4.2	8.8	2.9	64.7
8.6	18.3	14.4	4.3	6.2	10.6	4.4	21.4
57.4	43.7	12.6	3.7	7.6	13.0	3.4	-
25.7	24.7	11.2	4.3	9.5	11.0	4.4	
-	24.0	6.6	16.3	4.3	27.5	74.4	4.6
	12.8	6.9	9.7	6.2	15.2	15.2	8.8

The Mean Selection Percentage per school (N = 14) for each Item has been indicated above. The figures in the first cell indicate the percentage who correctly identified the family to which the Grid exemplar(s) belonged. In each cell, the Mean has been shown above the Standard Deviation.

APPENDIX 5.2

A Summary of Pupils' Responses to the Interview Questions

Candidate	Definition of Alcohol	Definiton of Ester	Definiton of Acid	Comparison of NaOH, H_2O and CH_3OH	First Reaction Pair	Second Reaction Pair
1	Contains OH group	Contains COCH group	Contains COOH group	NaOH and CH_3OH are Hydroxides	They're different (no reason). Acid + Alcohol = Ester + Water isn't the same as the other one.	Different and difficult.
2	No response	Alcohol and Acid make it	No response	All have OH group, but one has a non-metal, the others have metals. Couldn't sum up.	Different. HCl is a strong acid.	One has two acids, the other has no acids, so they're different.
3	Something to drink	Sweet smell	Fairly dangerous	Contain Oxygen and Hydrogen. Named water and ethanol. The OH group didn't mean that they had to react alike.	Both have H (first reactants) and OH (second reactants), and give water. But is an ester the same as Sodium Chloride? No conclusion.	You can block off the edge bits. -C(=O)-H and -C(=O)-O aren't the same because one has H and one has O. So the reactions are different.
4	Talk about bonds	No response	Talk about hydrogen	All have hydrogen bonding. Two are neutral, one alkali.	Both have an acid, but the alcohol is not the same as NaOH because it's neutral. So they're different.	Difficult substances.
5	Has OH	Acid and Alcohol make it	Weak acid. Has =O and OH groups.	Nothing the same. Ionic, water, alcohol.	Acid and alcohol giving ester is like salt and water but the reactions are not the same.	The first is an addition condensation. The second is an addition polymer. Hard formulae.
6	C with OH	Acid bonded with alcohol	COOH group.	Each has O joined to H but they don't behave alike.	Bits joining from both. The same sort of thing.	Both bonding together and leaving a bit over. They're hard. The formulae are difficult.
7	Contains OH group	Contains -C(=O)-O group	Contains COOH group	They're alkali, water and alcohol. They have different pH's.	The first - acid + alkali gives salt + water. The second is different. There is a C=O in the product so it's an ester.	Different. No water in the second reaction.
8	Alkane with H taken off and OH put on	Formed between alcohol and acid	Organic compound with a COOH group	Different. NaOH is ionic, H_2O is polar covalent and alcohol non polar covalent.	Different. First is acid + alkali = salt + water. The second is formation of ester.	Different. First has two acids giving acid + H_2O. The other is two esters going to a larger ester + alcohol.
9	OH group	C(=O)-O group	COOH group	All have H and all have O so they are same type of substance	Same type of reaction; Acid + Hydroxide, Acid + Alcohol.	Not the same. One gives H_2O, a condensation. The other has no H_2O.
10	Hydrocarbon with OH	Hydrocarbon with O between two C's	Contains C=O and OH groups	All got H bonding to some extent. All different types of molecules.	More or less the same; both acids and alkalis.	Same type of reaction because of the double bonds for O's.
11	Contains OH group	Add acid and alcohol and eliminate water	Weak acid. COOH group.	All have O-H group, different pH's	Not sure. Guessed they were different.	Don't know.
12	Ends in OH group	Compound between acid and alcohol	Ends in COOH group	NaOH is ionic, then water, then covalent	Water is given off in both. HCl is a strong acid. The second has a weak acid.	Same kind of compound formed.

APPENDIX 5.2 (Continued)

Pupils Reasons for Rejecting Test Items

Item	Grid 1	Item	Grid 2	Item	Grid 3	Item	Grid 4. Acid Series
	Candidate 2		Candidate 4		Candidate 5		Candidate 3
2	No C.	2	= between C's. Missing =O.	2	Isn't a primary alcohol	3	It's an ether.
3	Didn't look right.	3	Not right.		and no -OH group.	4	Is C=C the same a C–C?
4	It is a muddle.	4	Not right.	3	It's an ether.	5	The OH isn't matched.
6	The OH comes off differently.	6	The O is joined to H.	4	It's an acid.	8	=O is missing.
7	The two carbons with a double	11	= between C's.	5	It's a secondary one.	9	Is an O missing?
	bond.			7	Not COH.	10	It's got two.
9	O is in the middle and there		Candidate 7	8	It's a tertiary one.	15	=O is missing.
	is no H.			10	It's an acid.	16	Too many H's.
12	It's a jumble.	1	It's got a =O with no H –	11	It's ionic.		
14	Two O's.		that's what to look for.	12	It's an ether.		Candidate 8
15	O is missing.	6	Has O-H.	13	Unsure.		
		5	It's got two ='s. The C's	14	No COH group.	1	It's got an extra OH.
	Candidate 6		are opposite.	15	It's like an ether.	2	Hasn't got an OH.
2	The Cr.		Candidate 11		Candidate 10		Candidate 10
3	Couldn't picture it.						
4	Couldn't picture it.	2	It has a double bond to C	1	It's a primary alcohol.	2	The O is bonded to a C.
6	The N.		not O.	2	It's a secondary one, you		
7	Couldn't picture it.				should look for primary		Grid 4. Alcohol Series
8	The Br and Cl.				ones.		
9	Too many O's.						Candidate 1
12	Too many ='s.						
14	The C_6H_5.					1	An alcohol doesn't have
15	Can you match the separated						a double bond.
	bits?					2	An alcohol doesn't have
							a double bond.
	Candidate 7					3	The O hasn't got an H.
						5	The double bond.
2	It's got Cr.					9	Just guessed.
3	Has no COOH group.					10	The =O.
4	Doesn't appear the same.					11	The =O.
						16	No OH.
	Candidate 9						
							Candidate 8
1	It's a carboxylic acid.						
2	It's got two ='s.					8	This is an alcohol.
4	It has a ring – acids					12	This is an alcohol.
	don't have rings.					15	This is an alcohol.
	Candidate 11						
7	Has no COOH.						
	Candidate 12						
2	Has Cr, no C=O. You						
	should look for COOH.						

CHAPTER 6

Areas of Difficulty in Inorganic Chemistry and the I.C.C.U.D. Hypothesis

The results quoted and discussed in this Chapter have been taken from studies carried out by Howe,[83] Duncan,[84,85] and Garforth[86-88] respectively. At the outset, it should be clearly understood that the examination of these results was neither intended nor expected to form a critical test of the I.C.C.U.D. hypothesis. An experiment can provide a critical test of an hypothesis only if it is potentially capable of demonstrating a refutation of that hypothesis, and this critical function will be achieved only by careful experimental design, identifying appropriate data and establishing a specific and complete set of experimental questions. It would therefore be quite unreasonable to expect that these three studies, each of which was designed to answer its own specific questions, should also, fortuitously, provide a critical test of the I.C.C.U.D. hypothesis. However, each study was concerned with a difficult area of Chemistry, where one would expect the I.C.C.U.D. hypothesis to be relevant, and in addition, each required a careful and systematic collection of data, relating to pupils' performances in selected tasks within these areas. It therefore seemed that an examination of the results could well - and very probably should - reveal patterns of performance or behaviour consistent with the I.C.C.U.D. hypothesis (if this were valid).

The three studies to be examined covered one or more of the topics of the writing, balancing and interpretation of equations, the writing of formulae (of inorganic compounds) and calculations involving the concept of the mole. Because a different test design was used in each study, the procedure for examining the results had also to vary from study to study, but, broadly speaking, the investigation of each set of results fell into two parts. First, information about apparent levels of relevant conceptual understanding was sought, and then pupils' performances in tasks of different information content were compared.

Both Howe and Duncan had required pupils to work through a series of questions, which were effectively increasingly complex versions of related tasks. This allowed a fairly straightforward examination to be made of pupils' success rate as the information content - or transient memory load - of these versions increased. In the Organic Chemistry case, the Combined Tests' results had provided an independent measure of Short Term Capacity for organic formulae. No such measure was available for the Inorganic Chemistry tasks of interest here, and so we could not specify in advance what sort of task could represent a critical information load. For this reason, the criterion adopted for "performance consistent with the I.C.C.U.D. hypothesis" was an accelerating decrease, or a sudden decrease, in success rate with increase in information content, given a low level of relevant conceptual understanding.

The section of Garforth's study relevant to the present examination was structured very differently, in that pupils were not required to answer increasingly complex questions. Rather, they were asked to choose a response from several given responses, and it happened that these differed considerably in their information content. Now, the I.C.C.U.D. hypothesis would seem to imply that pupils with a low conceptual understanding would tend to prefer a low information situation (other things being equal) so these results seemed to be of great potential interest.

The three sets of results will now be considered separately, and the extent to which they were consistent with the I.C.C.U.D. hypothesis will be discussed in the concluding Section of the Chapter.

6.1 Howe's Investigation of Chemical Formulae and Equations

In Scotland, by 1970, only the Alternative 'O' Grade Chemistry syllabus was examined by the S.E.D. This syllabus was designed to promote the understanding of chemical principles, rather than the "rote" learning favoured by the traditional syllabus.

The treatment of symbols, formulae and equations was delineated in Memorandum Number 7[(89)] which recommended (among other things);

> "the formulae of an ionic substance may be written indicating the charges on the ions"
>
> "unless it is immediately obvious, it is useful to add subscripts to the symbols to indicate the state of the substance concerned"
>
> "it must be left to the discretion of the teacher when to introduce these conventions".

In spite of this caution, many teachers seemed to make widespread use of complex formulae and equations.

It was against this background that Howe carried out an investigation (from 1970 to 1974) into the writing of formulae and equations and their use. He described the situation extant in 1970 as "a very confused one, with, on the one hand official publications apparently advocating complexity, and on the other, some teachers becoming increasingly disillusioned as they tried to teach their pupils". He reported a feeling among teachers that the complex material was conceptually beyond many SIII and SIV pupils (14 and 15 year olds), and that because of the complexity, straight recall was likely to be inefficient.

As a major part of his study, Howe constructed a Gagne type flow diagram, illustrating the sequence of learning steps, required by the Alternative Syllabus, involved in writing binary formulae. (The intention of the Alternative syllabus was that these steps would be logically connected in that prior steps were to act as explanations or principles for later steps in a sequence). The hierarchy was then divided into several learning units or "stages", and Howe reasoned that if pupils were learning in the logical manner assumed, they should show competence in the terminal skill of a stage if and only if they also showed competence in the related prior steps. Pupils who had only the terminal skill correct could, he argued, be considered to have rotely learned these skills.

The prior steps and terminal skills were incorporated into a 17 item test, for which the numbers of pupils giving certain combinations of responses to the steps within a stage were recorded. A further 9 items were added to investigate pupils' ability to use the mole concept, to write and interpret equations, and to perform simple calculations. Response sequences for some of these items were also recorded. This test was designed to allow:

(a) a determination of the extent to which pupils relied on rote learning rather than the logical or structured learning that was the aim of the Alternative syllabus, and

(b) the identification of the step(s) within a stage at which a significant drop in performance occurred.

The results which will be considered here were obtained when this test was administered to a representative sample of 513 SIII pupils (14 year olds) some weeks after they had completed the relevant sections of the Chemistry course. Before these results are discussed, we will outline the procedure that was used to infer pupils' levels of conceptual understanding from Howe's test results.

6.11 Information Redundancy and Conceptual Understanding

We have previously discussed (Section 5.31) the way in which conceptual understanding can effectively decrease memory load by declaring some information redundant (at least in the first instance), and the converse situation in which pupils with low conceptual understanding may treat as necessary essentially redundant information (such as the details of a C-H chain). It is also possible that a low level of conceptual understanding can result in a pupil's finding necessary information redundant in certain contexts, or for certain purposes. Now, a pupil may as a result of his low conceptual understanding, ignore necessary information - just as he may, at other times treat redundant or extraneous information as necessary - but this is not the same as finding the information redundant. In the latter case, information given for a particular purpose cannot function as intended because of the low level of conceptual understanding. In other words, the generic meaning (corresponding to the extant level of conceptual understanding) is inadequate for the required task, although it may be adequate for other purposes. For example, given the number of electrons possessed by an element, a pupil may be able to describe the arrangement of these electrons, but may be unable to use this information to deduce the type of compound formed by the element; within this context, the information would have been redundant. From this, we would infer a low level of conceptual understanding of Electron configuration.

It is worth noting that, although a pupil may decide to disregard necessary information which is redundant for him, he may equally store and attempt to operate on it. Consider the symbols $Ag^{+}_{(aq)}$. A pupil may know that "aq" stands for "aqueous", but be unable to appreciate the difference between Ag^{+} and $Ag^{+}_{(aq)}$; for such a pupil the symbol "aq" would constitute redundant information in many contexts. However, he may feel that it ought to tell him something, and so will store this information, increasing his memory load without increasing his probability of success in the required task. (In fact the I.C.C.U.D. hypothesis would predict that such an addition to the memory load would decrease his chances of success, particularly in a high information context). Thus, even if it is known that some of the necessary information given is effectively redundant, it must still be considered as contributing to the load imposed by the required task.

Because of the format of Howe's test, the extent to which necessary information was effectively redundant served as an efficient indicator of levels of conceptual understanding.

Where pupils have to answer a two step question, they will fall into one of the four categories shown in Fig. 6.1, depending on their sequence of responses.

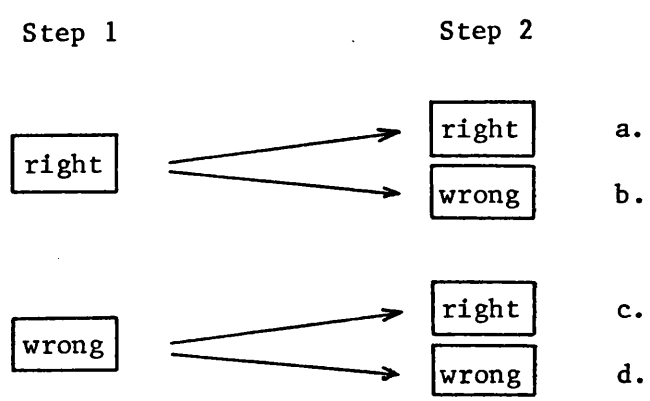


Figure 6.1 The Four Outcome Categories of a Two-step Question

Where Steps 1 and 2 have the relation prior step to terminal step, as in Howe's test, the pupils in category c (the wrong-right group) must have rote-learned the terminal skill. As mentioned above, Howe was particularly interested in identifying this group of pupils, for each of the learning stages. From our point of view, pupils in categories b and c were of interest, because for both these groups the information given in the prior step - step 1 - was clearly redundant in the context of the terminal skill (Step 2).

It should be noted that the status of the information of the prior step for pupils in groups a and d (the both right and both wrong groups) remains indeterminate, because it is impossible to distinguish between the cases

(a) both answers correct/incorrect due to correct/incorrect rote learning (when prior step information would be effectively redundant) and

(b) a logical progression from true/false premise to true/false conclusion.

Thus the total percentage of pupils in categories b and c represents the lower limit of pupils for whom Step 1 information is redundant in the context of Step 2. For a two-step stage, it would be sufficient to refer to this percentage as the Redundancy for that stage, but we chose to use the more explicit designation "(1-2) Redundancy"; this more formal system was essential when there were more than two steps per stage, if we were to specify exactly what information was redundant in which context.

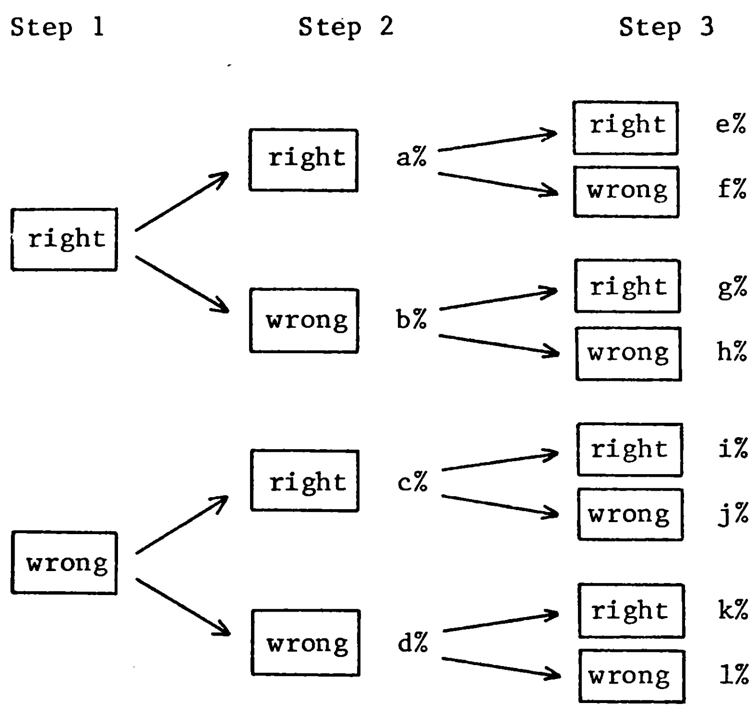


Figure 6.2 The Outcome Categories for a Three-step Question

Even in a three-step stage, as illustrated in Fig. 6.2, one could clearly compute several Redundancy figures. The simplest cases involve the computation of the percentage finding the information of one step redundant in the context of a succeeding step, neglecting the responses given for the third step. This gives rise to three Redundancies, which would be designated (1-2) Redundancy, (1-3) Redundancy, and (2-3) Redundancy. Referring to Fig. 6.2,

the (1-2) Redundancy = b% + c% (as before);

the (1-3) Redundancy = (f% + h%) (the right-wrong groups) +(i% + k%) (the wrong-right groups);

and the (2-3) Redundancy = (f% + j%) (the right-wrong groups) +(g% + k%) (the wrong-right groups).

One other case is worth mentioning - composite Redundancies, as for example, the percentage who find the combined information of Steps 1 and 2 redundant in the context of Step 3. This would be designated the

((1 + 2) - 3) Redundancy, and from Fig. 6.2 would be f% (right-wrong) + k% (wrong-right).

It was possible to calculate Redundancies within many of the stages from Howe's figures. In using these figures to infer levels of conceptual understanding some critical value of Redundancy had to be adopted. We were not concerned here with making fine distinctions between levels of understanding, but rather with identifying those instances for which the "overall" lack of conceptual understanding was sufficient for the decrement in performance predicted by the I.C.C.U.D. hypothesis to be observable, given the validity of the hypothesis. Bearing in mind that the Redundancies calculated were lower limits, we chose a value of 30% as the criterion; thus, where Redundancies exceeded 30% we inferred a related low level of conceptual understanding that was sufficiently widespread to be of practical consequence.

6.12 Levels of Conceptual Understanding

A summarised version of the Test Items 3-26, separated into stages, has been included in Appendix 6.1, together with two sets of data taken directly from Howe's results, namely the percentage of pupils correct for each step and the percentage who were correct for that step and all the prior steps within the stage. (In one or two cases, the cumulative results related to only some of the prior steps, and in these cases the relevant steps have been explicitly designated). The Redundancies computed within each stage have also been included. In some cases, Howe quoted figures for pupils having a "correct method" in the terminal step; where those figures have been used, an appropriate indication has been made. Where pupils were asked to respond to more than one example in a step, the number of correct responses which Howe required for competence has been indicated, and the "% correct" in these cases is in fact, the "% showing competence".

The results for questions (3)-(16) and (18)-(22) will be presented and discussed in this Section. For simplicity, the questions have been divided into four sets, each of which was concerned with a particular skill or concept. The questions forming each set are indicated in the text, and the three sets of numerical data already described - the percentage correct, the cumulative percentage correct, and the Redundancies - are given in Columns I-III respectively.

6.121 Questions Involving Electron Arrangements

Electron arrangements formed a prior step in the following items:

			I		II		III
(3)	Choose electron arrangement for $^{16}_{8}O$		73				
(4)	Oxygen completes shell by (a) {sharing} or (b) {gaining} electrons?	4a	62	(+3)	49	(3-4a)	36
		4b	71	(+3)	53	(3-4b)	37
		4a+4b	43	(+3)	33	(3-(4a+b)	50
(5a)	State electron arrangements (3/4)		76				
(5b)	State valence numbers (3/4)		66		57	(5a-5b)	29
(11)	Write electron arrangement for Calcium		78				
(12)	How many electrons will it lose to have a completely filled outer shell?		81		73	(11-12)	13
(13)	Write the symbol for the calcium ion		51		42	(11-13)	43
In addition, electron loss/gain formed a prior step in determining the symbol of an ion:							
(9)	Cl + (9a) .... Cl$^{(9b)}$	9a	69				
		9b	48	(+9a)	41	(9a-9b)	35
(12)	How many electrons will calcium lose to have a completely filled outer shell?		81				
(13)	Write the symbol for the calcium ion		51	(+12)	46	(12-13)	41

(Pupils were asked to give the electron arrangement of Cl (a prior step to (9a)), but their performance was not recorded).

It is instructive to compare the percentages who found electron arrangements redundant information in the different contexts. Questions (4) and (12) were, in principle, equivalent, but the different wordings gave very different percentages for Redundancy. The percentages of pupils who had the electron arrangement wrong but the second step right were comparable in the two cases, but the percentages going from right to wrong were very different (40% in questions (3-(4a+b)), and 5% in questions (11-12)). This pattern of responses suggests strongly that in a numerical context, an electron arrangement provided non-redundant (i.e. usable) information to many pupils, but in a more chemical context, it was redundant to many pupils.

This suggestion is further strengthened by observing the percentages who found electron arrangements redundant in stating valence numbers (5b) and the formula of the calcium ion (13), and in the redundancies shown in relating electron gain or loss to ion formulae in questions (9a,9b) and (12,13). (The results for questions 5a and 5b were somewhat ambiguous, because of the 3/4 correct competency criterion. If one were to assume that "both 5a and 5b correct" implied any 3 electron arrangements + any 3 valence numbers correct, rather than 3/4 pairs correct, the actual (5a-5b) Redundancy could well exceed the calculated value of 29%).

With the exception of the numerical context of question 12, the Redundancies within these learning stages were equal to, or greater than, the criterion value of 30%. In general then, it would seem that electron arrangements had some generic meaning for most pupils, but that for many, this did not allow the information to function as intended in a chemical context.

6.122 Questions Involving the Writing of Formulae

(a) Covalent Formulae

			I	II	III
(5a)	State electron arrangement	(3/4)	76		
(5b)	State valence number	(3/4)	66	57	(5a-5b) 29
(6)	Write formulae for four compounds, containing elements from question 5	(3/4)	43	35	
	More than 1 wrong, but correct method		20	?	
(7)	Name elements in four compounds	(3/4)	70		
(8)	Write formulae of these compounds (using the periodic table)	(3/4)	33	31	(7-8) 42
	More than 1 wrong, but correct method		19	13	

In questions (5,6) it was possible to obtain only an estimate of the (5a-6) and (5b-6) Redundancies. The results given by Howe indicated a non-trivial (5b-6) Redundancy of some 20-30%, and also a (5a-6) Redundancy exceeding 40%. There were other difficulties in calculating Redundancies for these questions. The uncertainty introduced by the matching of 3/4 correct in questions (5a,5b) obviously extended into question 6; also, it was not possible to account for the 20% who were assigned "correct method". Howe defined "correct method" as "incorrect because of slips, e.g. wrong symbols, wrong additions, etc." The elements in this question were carbon, hydrogen, oxygen, and fluorine, and their symbols were given; it is not easy to see what sort of mistake would be represented by "correct method". For the purpose of the Redundancy calculation, a trivial mistake would have been irrelevant, and such pupils included in the "correct" categories for question 6, but as their sequence of prior responses was not reported, their contribution to the Redundancies could not be determined.

The figures for the "correct method" group in question (8) were not used in computing the (7-8) Redundancy, again because of some uncertainty in the exact meaning of "correct method". In considering questions (7) and (8), it is important to realize that they formed the first and third steps of a three step sequence in which the responses to the middle step - deduce valency from the Periodic Table - were not recorded separately. Therefore, the computed (7-8) Redundancy of 42% actually represents those for whom either the element names were redundant for determining valency, or the valencies were redundant for writing formulae, but not both. As in the case of Questions (5) and (6), this (7-8) Redundancy could have been underestimated because of the uncertainty introduced by the two competency criteria.

The most serious consequence of the lack of preciseness in the Redundancy figures for Questions (5)-(8) was the impossibility of an accurate measure of the redundancy of valence numbers in determining a covalent formula. However, the figures that could be computed certainly suggested that a significant percentage of pupils were unable to make use of electronic configurations and/or valence numbers in writing covalent formulae, even though the relative magnitudes of the two Redundancies could not be determined.

(b) Ionic Formulae

While Questions (9)-(17) involved the writing of ionic formulae, only Questions (9)-(14) and (17e) will be considered here.

		I	II	III	
(9a)	Electron gain	69			
(9b)	Formula of chloride ion	48	41	(9a-9b)	35
(10)	Formula of sodium chloride	80	36	(9a-10) (9b-10)	31 46
(11)	Electron arrangement of Calcium	78			
(12)	Number of electrons lost	81	73	(11-12)	13
(13)	Formula of Calcium ion	51	42	(11-13) (12-13)	43 41
(14)	Formula of Calcium chloride	69	36	(12-14) (13-14)	30 36
(17e)	Formula for Sodium silicate	4			

The Redundancies in Questions (9a,9b) and (11-13), the prior steps for Questions (10) and (14) respectively, have been considered already.

The high Redundancies between electron gain/loss and compound formula - (9a-10) and (12-14), and between ion symbols and compound formula - (9b-10) and (13-14), suggest a low level of conceptual understanding. The fact that only 4% of the sample were able to deduce the formula of Sodium silicate suggested a very low level indeed.

6.123 <u>Questions Involving the Mole</u>

		I	II	III	
(19)	How many moles (gm. atoms) of Iron in 1 mole of Fe_2O_3?	32			
(18)	Give the formula weight of four compounds 3/4 correct <u>or</u> right method	81	29	(19-18)	58

The results of Question (19) suggest a low level of conceptual understanding of the mole. A comparison of the results of Questions (18) and (19) suggest that many pupils could have been advantaged by the use of a taught strategy for determining formula weights. The results quoted for Questions (20)-(22) (listed in Appendix 6.1) are also consistent with a low level of conceptual understanding of the mole. Howe reported that many pupils, in answering Question (20), showed an inability to determine when a symbol or formula represented some of a substance, as opposed to a mole of the substance.

In summary, the examination of the first three sets of questions revealed the following cases of non-trivial redundancy:

1. Electron arrangements in the context of:
   (a) electron transfer or sharing to achieve stability
   (b) valence number
   (c) ion formulae.
2. Electron gain or loss in the context of:
   (a) ion symbols
   (b) ion ic formula writing.
3. Valence and/or electron arrangement in the context of covalent formula writing.
4. Ion symbols in the context of ionic formula writing.

From these we inferred a low level of conceptual understanding of the fundamental chemical principles of electron configuration, electron transfer or sharing and valence. This could perhaps be expressed more succinctly as a lack of conceptual understanding of

(i) ions as species,

and (ii) the formation and composition of simple binary compounds.

In addition, a low level of conceptual understanding of the mole was inferred from the results of Questions (18)-(22).

In a general way, the lack of conceptual understanding mentioned above would be expected to cause difficulty in the balancing and interpreting of equations, and the solution of simple mole calculations, which were the variable information areas investigated by Howe. Pupils' performance in this part of his test will be considered in the next Section.

6.13 The Relation Between Performance and Information Content

The results of Questions (15)-(17) and (23)-(26) were the main areas in which a comparison of performance in different information contexts was made. These questions have also been divided into four sets, each concerned with a particular concept or technique. As in the last Section, summaries of these questions have been reproduced in the text, and the percentage of pupils correct, or demonstrating competency, has also been indicated.

6.131 Ionic Formulae

			% Correct
(15)	Given the names and charges of six ions, write formulae (names and charges) for four compounds.	(3/4)	63
(16)	Give the formulae (symbols and charges) of 5 oxyanions.	(3/5)	50
(17a-d)	Write the formulae of four compounds containing these anions.	(3/4)	27

In Question (15) pupils were instructed to deduce the formulae by balancing charges, and were given a worked example. This could have acted as a taught strategy, which would have concentrated attention on a small amount of information.

Only 24% of the sample were reported correct for both Question (16) and Questions (17a-d). This suggests that many pupils were unable to apply the same rule in the higher information context of Questions (17a-d). The drop in performance from Question (15) to Questions (17a-d) is consistent with the I.C.C.U.D. hypothesis.

6.132 Mole and Weight Calculations

		% Correct
(23b)	Given: $2Pb(NO_3)_2 \rightarrow 2PbO + 4NO_2 + O_2$; How many moles of O_2 from 1 mole of $Pb(NO_3)_2$?	3
(24)	Given: $CaCO_3 \rightarrow CaO + CO_2$ F.Wt. 100 56 44; What weight of CO_2 is made by completely roasting 15 gms of chalk?	30

As the formula weights in Question (24) were given, pupils who could equate "chalk" and "$CaCO_3$" could adopt a strategy of considering given numbers only. That is, they could succeed by considering only a limited amount of information, independent of their conceptual understanding of the mole. The numerical computation required in Question (23b) was much simpler than that in Question (24), but conceptual understanding was required in handling the total information provided.

Thus, in comparing the two Questions, the one in which pupils' attention was drawn to a small amount of information, and which was susceptible to a strategy based on "student cunning" rather than conceptual understanding, had a success rate an order of magnitude greater than the other.

It could be argued that the lack of success in Question (23b) was due to the effective redundancy of necessary information (consequent upon lack of conceptual understanding) and was independent of information content. However, we do observe a success rate of 32% in the low information question "How many moles of iron are there in 1 mole of Fe_2O_3"(Question (19); the order of magnitude drop in success rate suggests that performance was information content dependent. The comparison of moles "within" a compound ($Fe_2O_3 \rightarrow 2Fe$) and moles "of" a compound is not the best; this type of calculation was examined in more detail in the next investigation and further consideration of it will be deferred until Section 6.221.

It might also be expected that under the conditions mentioned, more pupils should have been successful in Question (24). Two points should be mentioned here:

(i) The results of Questions (18) and (19) suggested that many pupils adopted a simple strategy - perhaps purely mathematical - in answering Question (18). Question (24) represents a reasonably large increase in information content, and required a more complicated strategy, so the considerably lower success rate in Question (24) is to be expected, given the I.C.C.U.D. hypothesis.

(ii) Given the apparent low level of conceptual understanding of the mole and compound composition, it would require some sophistication on the part of a pupil to ignore the chemical information, and adopt a "mathematical" strategy. Thus, a low success rate would be expected. (Support for point (ii) was found in an interview conducted by Urquhart[(90)] as part of her investigation into pupils' ability to solve proportion problems in chemical contexts. In solving chemical proportion problems, it was observed that most pupils who said they "didn't really understand the problem" simply did not attempt to solve it, but some said they "just tried with the numbers"). It should be noted that Howe gave a different interpretation of the results of Questions (23b) and (24). He reasoned that those who had (23b) wrong, but (24) correct - approximately 28% of the sample - failed Question (23b) because of "chemical difficulties". He stated that the remaining 72% therefore failed Question (24) because they were unable to do simple proportion. We would suggest that the mere presence of chemical information - even though it is rendered largely redundant by the form of the question - can prevent success where there is a low level of conceptual understanding. (It is of interest to note that Urquhart's results indicated that some pupils who were successful in "mathematical" proportion failed in "chemical" proportion). From the point of view of the I.C.C.U.D. hypothesis, conceptual understanding of both relevant chemical principles and proportion would be required for a high success rate in a chemical proportion problem, where the information content was high.

6.133 Balancing Equations

Balance:		% Correct
(25a)	$Ca(OH)_2 + HCl \rightarrow CaCl_2 + H_2O$	27
(25b)	$Ca^{2+}_{(aq)} + 2Cl^-_{(aq)} + Ag^+_{(aq)} + NO^-_{3(aq)} \rightarrow$	
	$Ca^{2+}_{(aq)} + 2NO^-_{3(aq)} + Ag^+Cl^-_{(s)}$	32
	Both correct	18

Although there were no lower information questions to which these results could be compared, an interesting comparison between the questions was possible. Question (25b) appears to contain more information than

Question (25a), but the percentage giving a correct response to Question (25b) borders on being significantly greater than the percentage for Question (25a) (w_{max} ~ 6% at $\alpha = .05$). At first sight, this would seem to be a counter example to the I.C.C.U.D. hypothesis. However, in balancing an equation, an holistic consideration of information is not necessarily required - in fact, a "step by step" strategy is perfectly appropriate. It may be that when the ions are separated, pupils with low conceptual understanding actually find the problem easier, as they are encouraged to concentrate on one species at a time. Thus, these results are not inconsistent with the I.C.C.U.D. hypothesis.

6.134 Interpreting Equations

Rewrite the following sentences, using only words:

		% Correct
(26a)	He sprinkled NaCl into a beaker full of H_2O	81
(26b)	$H_2O_{(1)}$ and $Mg_{(s)}$ do not react readily to give $H_{2(g)}$, but when $H_2O_{(g)}$ is passed over hot $Mg_{(s)}$ it gives $H_{2(g)}$ and $MgO_{(s)}$.	32
(26c)	$Ag^+_{(aq)} + NO^-_{3(aq)}$ were added to $H^+_{(aq)} + Cl^-_{(aq)}$ to give $Ag^+Cl^-_{(s)} + H^+_{(aq)} + NO^-_{3(aq)}$	5

These questions required an holistic or converging treatment of the information presented. As the information - and particularly the "chemical" information - increased, a dramatic drop in success rate occurred. The symbols used in the three examples were similar in type, although not identical, so this would seem to be an information content effect. This view is supported by a comparison of Questions (25b) and (26c). The equations involved were very similar, but when an holistic interpretation was required there was an order of magnitude drop in performance.

These results are consistent with the I.C.C.U.D. hypothesis. However, it is possible that effective redundancy of some of the symbols contributed to the low success rate, independent of the information content.

6.14 Difficulty Ratings

As a separate part of his study, Howe asked pupils to rate each of several principles as "Very easy", "Easy", "Hard", or "Very hard". He assigned weightings of 1-4 (Very easy → Very hard) to these categories, and multiplied the weighted averages for each principle or technique by 20 to obtain a "difficulty indicator" of the range 20-80. Although this method of analysis tends to be insensitive to moderate or small differences in difficulty designation, it does give an indication of pupils' difficulty ratings.

The difficulty indicator for the principles or techniques that formed the basis of Howe's test have been reproduced in Table 6.1. These ratings appeared to fall naturally into the three categories indicated. The I.C.C.U.D. hypothesis would suggest that high difficulty ratings would be given to "high information" techniques, while medium or low ratings would be given to "low information" techniques, or "medium information" techniques were a strategy could be employed with some success (independent of conceptual understanding). Allowing that rote learning could be termed a strategy, the results quoted in Table 6.1 are consistent with the I.C.C.U.D. hypothesis.

TABLE 6.1

DIFFICULTY RATINGS

Technique	Rating	
Writing formulae for simple binary compounds. The use of valence numbers.	39.6	LOW
Valence numbers from the Periodic Table.	42	
Writing word equations.	45	MEDIUM
Writing formulae for compounds such as Calcium carbonate.	46	
Putting the charges on formulae.	49	
Calculation of formula weights.	50	
Writing symbol equations.	54	HIGH
Balancing equations.	55	
Writing equations putting in state symbols.	59	
Writing equations with separate ions.	60	
Mole and calculations using it.	60	

6.15 Summary of the Examination of Howe's Results

A low level of conceptual understanding relevant to the writing of formulae, the balancing and interpreting of equations, and calculations involving the mole was inferred from Howe's results. A comparison of success rates with increasing information content, in different contexts, showed that the results were consistent with the I.C.C.U.D. hypothesis. In the cases of mole calculations from equations and interpreting equations the results did not exclude the possibility of an information content independent effect contributing to the drop in performance.

6.2 Duncan's Investigation of the Mole, Molarity, and Simple Volumetric Calculations

One of the areas of difficulty reported in Johnstone's survey[9] was the mole and its use in calculations from equations. In 1972-1974 Duncan carried out an investigation of pupils' performance in this area of Chemistry. His study had three aims:

(i) To determine as precisely as possible where difficulties occurred in using the mole and related concepts.

(ii) To compare the results obtained following different presentations of the topic. (Duncan suggested that this topic might not be intrinsically difficult, but might have become difficult for pupils because of certain teaching methods).

(iii) To determine the extent to which the difficulties in this area were a problem of maturity.

It was primarily the first part of Duncan's study (that is, the part concerned with aim (i)) that was pertinent to the present investigation, because it included a very detailed, almost step-by-step, examination of pupils' performance in tasks ranging from the determination of mole weights to simple volumetric calculations. Because pupils were often required to perform computations in Duncan's tests, we will preface our examination of his results with a discussion of the way in which computations affect the relation between information content, conceptual understanding and performance.

6.21 Computations, Information Content, and Conceptual Understanding

Even where conceptual understanding can reduce the total information content of a given question to a comprehensibly small number of chunks, the required answer (to, say a simple volumetric calculation) will not then appear because, generally speaking, a computation will be necessary. Thus, a step-wise strategy must generally be employed, whatever the information content of the question.

In areas considered previously, (such as balancing and interpreting equations) a strategy for dealing with a large block of information was not taught explicitly, so any strategy or chunking employed could be seen as a consequence of, and a reflection of, conceptual understanding. However, in problems that include computational steps, pupils are generally taught a solution strategy or rule, so the use of a strategy does not necessarily directly reflect their conceptual understanding. In other words, the taught strategy could function as an organisational aid, and so be confounded with conceptual understanding. It would seem certain that in low or medium information contexts, a pupil with low conceptual understanding would be advantaged by the use of a taught strategy, rather than a strategy based on his understanding (or misunderstanding). Compare, for example, the success rates in formulae weight calculations (using a taught strategy) and in organic family categorisations (using a devised strategy). However, as the following analysis will show, this advantage diminishes in a high information context, where the pupil is again heavily dependent on his own conceptual understanding.

First, the rule itself must be remembered. In a low information context, a simple mechanical rule can be taught which will cover all possibilities, enabling success (in principle) to be independent of conceptual understanding; but as the information content is increased, such a mechanical rule would become complex to the point of unmanageability. Thus the rule itself will tend to be expressed in a generalised or abstracted way; the generic meaning assigned to the rule will depend therefore on the level of conceptual understanding of relevant chemical principles. If we believe with Ausubel that meaningfully learned material is retained better than rotely learned material, conceptual understanding is advantageous in remembering the rule.

Secondly, when a given problem is read by a pupil, he must at least extract sufficient information from the text to identify the type of problem. His ability to do this will depend on his level of conceptual understanding; a taught solution strategy would not assist in this process. (At this stage, as we have already mentioned in Section 6.11, lack of conceptual understanding may result in the declaration of redundant information as necessary, needlessly increasing memory load and possibly also causing a pupil to abandon a problem before even attempting a solution. In the interview conducted by Urquhart (quoted earlier) some pupils said they could not solve a problem because they did not know the meaning of "neutralise" - a quite extraneous term in the particular context).

It is at the third stage - the performance of the computational steps - that we might expect the relation between information content, conceptual understanding and performance to be attenuated, given a taught solution strategy. To clarify this relationship, we will consider two strategies for solving a simple volumetric calculation. (These particular examples were chosen because they were used by Duncan in the second part of his investigation, and, as will be seen, the I.C.C.U.D. hypothesis makes a prediction of the comparative success rate to be expected for them). The two taught solution rules, and a possible verbalisation of each, are given below for the following problem type:

What volume of 2M "A" solution would completely neutralise 100 ml of 3M "B" solution?

Given: $A + 2B \rightarrow C + D$

<u>Solution Strategy 1</u>

From the balanced equation:

1. 1 mole B will neutralise .5 moles A

2. $^{1}/10$ litre of 3M B contains .3 moles B

3. .3 moles B neutralise .15 moles A

4. Volume of 2M A containing .15 moles A

   = (no. moles/molarity)

   $= \frac{.15}{2}$

   = .075 litre

   . 75 mls

<u>Verbal Representation</u>

1. Compute the number of moles of one substance reacting with 1 mole of the other.

2. Compute the number of moles of the substance for which volume and conc. are given.

3. Determine the corresponding number of moles of the other substance.

4. Determine the required property.

<u>Solution Strategy 2</u>

From the balanced equation:

1. 1 mole B will neutralise .5 moles A

2. 1 litre 1M B neutralises 1 litre .5M A

<u>Verbal Representation</u>

1. Compute the number of moles of one substance reacting with 1 mole of the other.

2. Write the molarity of one solution such that 1 litre of it will neutralise 1 litre of 1M other substance.

3. 1 litre 3M B neutralises 1 litre 1.5M A	3. Use the molarity of the solution for which volume and molarity are given, and deduce the molarity of the second substance such that 1 litre of it will react with 1 litre of the first solution.
4. $\frac{2}{1.5}$ litre 3M B neutralises 1 litre 2M A	4. Convert to the molarity of the second solution, and deduce the volume of the first that reacts with 1 litre of the second solution.
5. $\frac{1}{10}$ litre 3M B neutralises $\underline{x}$ litre 2M A $x = \frac{1/10}{2/1.5}$ $= .075$ litre $= 75$ mls	5. Use the given volume of the first solution and deduce the required property.

For someone with good conceptual understanding, a rule expressing either taught solution strategy is of course unnecessary - the information given in the text of the problem "speaks for itself". A rule has to supply explicitly for pupils with low conceptual understanding facts and particularly relations which are implicit in the given textual information to pupils with good conceptual understanding. A pupil who is dependent on the rule must store the "rule" information, as well as the textual information; he must therefore require more chunks than the pupil with good conceptual understanding, who can express the total in a few (textual + implicit) chunks. So, although the rule provides, for the pupil with low conceptual understanding, information without which he could not solve the problem, it is a pupil's level of conceptual understanding that will determine the number of chunks that will represent, for him, the total (given and inferrable) information necessary for the solution of the problem. Given that the total number of chunks depends on conceptual understanding, it remains to be seen whether the number of chunks that must be considered at any one time during the different phases of problem solution, depends on conceptual understanding, or on the taught strategy.

If a pupil is to handle problems of this type confidently, and with a high success rate, it would seem necessary to be able to take an overall view of the problem and the solution strategy. For a pupil constrained by his lack of conceptual understanding to use a taught rule, this would require that he should be able to comprehend the rule "in a one". He would presumably require at least one chunk to represent each rule step; in this respect, the rule may specify the minimum number of chunks required. However, depending on his precise level of conceptual understanding, he may require more than one chunk for some step(s). Possibly a generalised rule step may have to be replaced by specific alternatives (e.g. step 4 in strategy 1 might have to be stored as alternative rules for determining molarity and volume). From this it follows that the actual number of chunks required to represent a pupil's version of the taught strategy rule will depend largely on his level of conceptual understanding; if the minimum number of chunks required is close to the storage capacity of Working Memory, conceptual understanding will be very important in determining whether or not the required information can be considered simultaneously. Strategies 1 and 2 considered above would require a minimum of 4 or 5 chunks; these chunks must also be stored as an ordered sequence, which places a further burden on storage capacity.[(78)] It would seem, then, that the information content of the rules of strategies 1 and 2 comes dangerously close to the limit of Working Memory storage capacity, and for many pupils with low conceptual understanding, Working Memory capacity could well be exceeded.

Before performing a computational step, the relevant information must be extracted from the text. A pupil using a rule would need to keep at least the relevant step of the rule in Working Memory, and scan the problem text for the necessary information. (At this stage, the pupil may need to store a more detailed version of the rule step than was necessary for the overview of the problem and its solution). The rule will determine the amount of information that must be extracted: in the cases of strategies 1 and 2, this amount would seem to be small. However, lack of conceptual understanding may well cause redundant (for that step) information to be considered. Also, the number of chunks required to represent the rule step will depend on conceptual understanding. So once more, the level of conceptual understanding will be very important in determining whether or not storage capacity will be exceeded.

While the pupil is performing a computational step, the amount of information which <u>must</u> be handled is determined by the taught strategy. If this amount of information is large, conceptual understanding will determine whether or not chunk capacity is exceeded. (Again, lack of conceptual understanding may result in additional memory loading of either effectively redundant or extraneous information). The computational steps designated by strategy 1 are independent, and each requires only a moderate amount of information to be handled simultaneously. (The largest amount is involved in "if 1 mole of A gives <u>X</u> moles of B, y moles of A give ... moles B"). However, strategy 2 requires the use of large amounts of information (the volume and molarity of solutions in both the current and preceding steps, and also, at times, some of the given values of volume and molarity). In this case, conceptual understanding will be very important in deciding whether or not storage capacity will be exceeded.

To summarise then, a taught strategy may be expected to improve performance in a low information context, but in a high information situation,

(i) The number of chunks that can represent the total necessary information depends on conceptual understanding.

(ii) The taught strategy determines the minimum number of chunks that could represent the strategy rule for someone with low conceptual understanding. As these chunks must be ordered, even a small minimum number will approach Working Memory capacity. Therefore, conceptual understanding will be very important in determining whether or not an overview of the problem and its solution is possible.

(iii) The strategy will determine the amount of necessary information to be extracted from the text at any one time. However, the actual amount of information considered, and the number of chunks required to represent this, will depend on conceptual understanding.

(iv) The amount of information to be handled during any computational step is determined by the strategy. Where this amount is large, success will again depend on conceptual understanding.

Thus, from the point of view of the I.C.C.U.D. hypothesis, we would expect the predicted drop in performance associated with increase in information content to be lessened when a solution strategy has been taught, but we would still expect a noticeable decrement in performance, particularly when situations (iii) and (iv) above were high information contexts.

In considering volumetric calculations in particular, the I.C.C.U.D. hypothesis would predict that pupils with low conceptual understanding (particularly of the mole) would perform poorly whether they had been taught strategy 1 or strategy 2. In comparing these two strategies, the only major difference from the information point of view, occurs in the actual computations, where strategy 2 involves the use of a much larger quantity of information simultaneously. Therefore, the I.C.C.U.D. hypothesis would predict that pupils with low conceptual understanding who have been taught strategy 2 would perform even more poorly than those taught strategy 1.

It follows from the above discussion that, because of the improved performance to be expected in a low information context involving a taught solution strategy, such examples should not be used in determining the level of relevant conceptual understanding.

It would also seem that certain attributes of problems could lessen the effect of lack of conceptual understanding. Some computations may be particularly susceptible to a purely mathematical strategy. For example, pupils could use the strategy "multiply volume by numerical prefix of M" to compute the number of moles, independent of their level of conceptual understanding of molarity.

In some formats of volumetric calculations, the information about each substance is physically brought together, and clearly related to the substance. This may act as an organisational aid (by bringing about more efficient chunking, or by reducing the amount of redundant material considered while extracting information from the text), and hence lessen the effect of lack of conceptual understanding.

Thirdly, it is possible that the easily seen relation between simple numbers may act as an organisational aid, again reducing the effect of lack of conceptual understanding. Such a relation could effectively increase the information content of a chunk, or reduce the number of chunks required (e.g. "1 litre of 1M" may be representable by fewer chunks than "30 mls of 2.4M", or the latter may require consideration of additional rule instructions or an exemplar); it may also serve to link two quantitites whose chemical relation is only tenuously recognized.

It would seem essential that consideration should be given to all these factors in interpreting performances in computational tasks.

6.22 Discussion of Results

The set of results to be considered here came from a series of post-tests that were administered in 1973 to a representative sample of 500 SIII pupils (14-15 year olds). Half of this sample had worked through a series of four programmes designed by Duncan covering the mole, calculations involving the mole and simple volumetric calculations, and had completed a pre- and post-test for each learning program. The remainder of the sample had completed the same series of post-tests after class work covering the topics treated in each program. (The tests used for this experiment were a revised version of tests that had been trialled in 1972). In particular, we considered 29 objective items, taken from the second, third and fourth post-tests. A summary of these items is given in Appendix 6.2. Following Duncan, we have recorded the results of the Programme Group (P. Group) and the Classwork Group (C. Group) separately.

Duncan reported the Facility Value (F.V.) and the Discriminating Power (D.P.) for each item. (These give the proportion of a group choosing the correct response for an item, and the difference between the F.V. for the top and bottom thirds of the whole sample group, respectively). Because sequences of responses were not recorded, we could not apply the technique of computing information redundancy which had proved useful in the examination of Howe's results. However, we again considered groups of questions, each of which was concerned with a particular concept or technique. For each group, we

(i) examined the results for evidence of low levels of (relevant) conceptual understanding;

(ii) compared pupils' performance in related low and high information contexts.

As before, the questions considered within a group will be listed in the text, together with an indication of the type of response offered. Because of the detailed nature of Duncan's investigation, we were also able to

(iii) compare performance in isolated steps of volumetric calculations and in complete volumetric calculations;

(iv) investigate the extent to which the factors mentioned at the end of the previous Section appeared to lessen the effect of lack of conceptual understanding.

Finally, we compared the results obtained from questions common to Howe's and Duncan's tests. This comparison was valid, as the two had been drawn from the same population, at about the same time. It was of interest because Howe's tests were given some weeks after the related learning, whereas Duncan's were given immediately after learning.

6.221 Conceptual Understanding of the Mole, and Mole Calculations from Equations

Two sets of questions were considered here. The first group came from the first post-test and involved:

A. Calculation of Gram Formula weights.

B. Calculation of Mole weights.

C. Weight of integer or fractional numbers of moles.

D. Number of moles in a given weight.

Duncan reported F.V.'s of .8-.9 for these calculations, where formulae were given. The other items were:

		F.V.	
		P. Group	C. Group
(12)	How many moles of NaOH react with 1 mole H_2SO_4?	.58	.76
(13)	How many moles NO_2 from 1 mole $Pb(NO_3)_2$?	.46	.50
(14)	How many moles H_2 react with 1 mole N_2?	.41	.54
(15)	What weight Mg reacts with 32g S?	.66	.74
(17)	What weight O_2 reacts with 3g C?	.65	.77
(16)	What weight of SO_2 reacts with 32g O_2?	.24	.34
(18)	What weight of Al reacts with 80g CuO?	.27	.30

(An unbalanced equation was given for item (14), but the equations given for the other items were balanced).

As the calculations A, B, D, involved the use of taught strategies, there were unfortunately no low information questions that could be used to determine the level of conceptual understanding of the mole. However, for item (12), 15% of the C. Group and 30% of the P. Group chose 1 mole of NaOH, while for item (13) about 15% of both groups chose 1 mole of NO_2, and 30% of both groups chose 4 moles of NO_2. These results, along with the quoted F.V.'s for items (12)-(14) certainly suggested a low level of conceptual understanding of the mole.

It is difficult to determine the extent to which these results were affected by the information content of the questions. It seems, from the results for calculation C, that most pupils could interpret "the weight of 4 moles of X ..." as the "weight of 4 lots of X ...", and a similar interpretation of the equation in item (12) as "2 lots of NaOH + 1 lot of H_2SO_4" would be perfectly adequate. This would certainly suggest that the information content of items (12)-(14) contributed to the poor performance. It would indeed seem likely that the information in all but the simplest equations would exceed chunk capacity, if all the given information were considered. However, here as in the case of Howe's results, it is possible that effective redundancy of some given information made an independent contribution to the poor performance. If mole numbers were effectively redundant, we would expect the higher F.V.'s observed for items (15) and (17). However, this greater success rate could also be attributed to the lower information content of these items, particularly in the computational stages.

Items (16) and (18) were the highest information questions in this group. They can be considered as (Item (12) + Calculation C. + calculation D) - or possibly as a combination of (Item (12) + Item (15)). In either case, a dramatic drop in performance was observed, going from the individual steps to the combined calculation. It might be argued that, as F.V.'s and not individual sequences of responses were recorded, the F.V.'s for items (16) and (18) simply represented those who had all of C., D., and Item (12) - or Items (12) and (15) - correct. If we consider C., D., and Item (12) to be independent questions, we would expect F.V.'s for Items (16) and (18) of about .36-.48 (P. Group) and .48-.60 (C. Group). The alternative combination would give .40 (P. Group) and .54 (C. Group). (A F.V. can be considered as the probability of answering an item correctly. Thus, the probability of answering question Item (16) correctly would be F.V.(12) x F.V.(C.) x F.V.(D.) etc.) As the observed F.V.'s were much lower than these values (w_{max} ~ 9%, α = .05 for either of the groups) it is quite proper to relate the drop in performance to the increase in information content. This drop in performance is consistent with the I.C.C.U.D. hypothesis.

6.222 Conceptual Understanding of Molarity and Concentration

The examples to be considered were:

		F.V. P. Group	F.V. C. Group
(1)	A molar solution of HCl contains (given 4 definitions)	.28	.58
(2)	Which of the HCl solutions is most concentrated? (4 like 500 ml of 2M HCl)	.44	.50
(3)	Which solution of NaCl is most concentrated? (4 like 200 mls containing 2 moles NaCl)	.49	.57
(4)	If one mole of Sodium Hydroxide (NaOH) is dissolved in 500 ml of solution, what is its concentration?	.38	.56
(5)	If .5 moles of NaOH are dissolved in 200 ml of solution, what is its concentration?	.56	.77
(6)	Which solution contains most NaCl? (4 like 500 ml of 2M NaCl)	.48	.51
(8)	How many moles of NaOH are dissolved in 500 ml of 4M NaOH?	.64	.81

The different F.V.'s for Items (4) and (5) seemed anomalous, but overall the F.V.'s for the first 6 items suggest a low level of conceptual understanding of molarity and concentration. (It is possible that pupils could apply a purely mathematical strategy more successfully given ".5 moles of NaOH" - Item (5) - rather than "one mole of Sodium Hydroxide" as in Item (4)). In Item (6), Duncan reported that pupils were split almost equally between the largest volume (the correct response) and the largest concentration given. The results for Item (8) suggest that pupils may well have used a mathematical strategy for computing the number of moles, and so were reasonably successful in spite of their lack of conceptual understanding of molarity and concentration. The results given for Items (10) and (11), Appendix 6.2, would support this view. Thus, in examining performance in volumetric calculations, the use of such strategies must be considered.

6.223 Format and "Simple Number" Effects

Before considering the volumetric problems, it is also necessary to study the possible affect that different formats and the use of simple numbers might produce. The relevant items were:

(a) Simple Number Effect

		F.V.	
		P. Group	C. Group
(6)	Which solution contains most NaCl? (4 like 500 ml of 2M NaCl)	.48	.51
(7)	Which of the following solutions contains most NaCl? (4 like 30 ml of 1.2M NaCl)	.34	.35
(8)	How many moles of NaOH are dissolved in 500 ml of 4M NaOH?	.64	.81
(9)	How many moles of H_2SO_4 are dissolved in 15 ml of 2M H_2SO_4?	.32	.58

	F.V. P. Group	F.V. C. Group
(19) How many moles Mg react with 1 litre 1M H_2SO_4? (given a balanced equation)	.89	.90
(20) How many moles Mg react with 100 ml 4M H_2SO_4? (given a balanced equation)	.40	.78

In each of the pairs quoted above, the only observable difference was the type of numbers used. The use of numbers having an easily seen relation (Duncan described them as "easily imagined" numbers) certainly seemed to increase the success rate. It should be noted that Items (23) and (24) appear to be a counter example.

(b) Format Effect

	F.V. P. Group	F.V. C. Group
(23) What volume of 1M NaOH reacts with 2 litres of 1M HCl solution?	.69	.81
(25a) ½ litre of 1M NaOH is neutralised by 1 litre of HCl solution. What is its molarity?	.50	.61
(26) What volume of 1M NaOH will neutralise 1 litre of 1M H_2SO_4?	.36	.58
(28) 1 litre of 1M NaOH neutralises ½ litre of H_2SO_4. What is its molarity?	.22	.27

(A balanced equation was given for each Item).

These two pairs of Items allow a comparison of the formats "what volume of <u>x</u>M A reacts with ..." and "... react with <u>z</u> litres B. What is the molarity of b". (Other questions which showed the same format difference were not considered because they also involved the simple number effect). These results indicated that the format of a problem may affect the success rate.

6.224 Volumetric Calculations

In comparing the performances associated with isolated calculational steps and combined calculations within a volumetric problem, or between volumetric problems where the total information increases, it is important not to confound the simple number effect or a format change with a change in information content.

Given this constraint, the following comparisons were made:

	F.V. P. Group	F.V. C. Group
(8) How many moles of NaOH are dissolved in 500 ml of 4M NaOH?	.64	.81
(12) How many moles of NaOH react with 1 mole of H_2SO_4?	.58	.76
(8) and (12) are isolated steps for:		
(27) What volume of 2M H_2SO_4 will neutralise 250 ml of 4M NaOH?	.24	.23
(12) is an isolated step for:		
(26) What volume of 1M NaOH will neutralise 1 litre of 1M H_2SO_4?	.36	.58
(19) How many moles Mg react with 1 litre of 1M H_2SO_4?	.89	.90
(23) What volume of 1M NaOH reacts with 2 litres of 1M HCl?	.69	.81

(Balanced equations were given for all but Item (8)).

The three comparisons show drops in performance consistent with the I.C.C.U.D. hypothesis.

6.225 Strategies 1 and 2: A Comparison of Performance

The fourth learning programme, which covered volumetric calculations, was produced in two versions, one using strategy 1 and the other strategy 2. The mean scores for the two groups for pre- and post-test Items (Items (19)-(29)) were recorded, and are reproduced here.

		Strategy 1		Strategy 2
Pre-test	(N=85)	5.3	(N=145)	4.5
Post-test	(N=86)	6.3	(N=140)	4.7

Clearly, neither group scored well on the post-test, which agrees with the I.C.C.U.D. hypothesis. The group using strategy 2 showed no improvement at all; Duncan reported a difference significant at $\alpha = .01$ for the combined programme group, and it therefore follows that the group using strategy 1 must have made a significant improvement. This suggests that the group taught strategy 2 were disadvantaged compared to the group who were taught strategy 1, which, again, is consistent with the I.C.C.U.D. hypothesis.

6.226 A Comparison of Immediate and Delayed Test Results

The I.C.C.U.D. hypothesis predicts that, given a lack of conceptual understanding, performance will drop dramatically as information content increases; also, lack of conceptual understanding would be expected to reduce performance in a delayed test. Putting these together, we would expect the drop in performance from immediate to delayed tests to increase with the total information content of the test task, given that the hypothesis is valid. The results of the questions common to both experiments are given in Table 6.2.

TABLE 6.2

RESULTS OF IMMEDIATE AND DELAYED TESTS

Question	Immediate Test Mean F.V.	Delayed Test F.V.
Calculation of Formula weight	.85	.81
Balancing an equation	.40	.27
How many moles of A from 1 mole B:		
(a) Numbers read directly from equation	.60	.20
(b) Numbers computed	.48	.03
Calculation of weight of A obtained from given weight of B (1:1 mole ratio)	.70	.30

These results are also consistent with the I.C.C.U.D. hypothesis.

Overall, the examination of Duncan's results revealed a low level of conceptual understanding of the mole, molarity and concentration. We had suggested that the dependence of success in volumetric calculations on conceptual understanding of the mole molarity and concentration could be lessened by the use of a mathematical or numerical strategy; Duncan's results contained examples consistent with the use of such a strategy. Other examples showed that the simple number effect and a particular format appeared to increase the success rate. (One counter example was noted).

In weight from weight calculations, and in volumetric calculations drops in performance were observed which were consistent with the I.C.C.U.D. hypothesis. In many cases these drops were large, in spite of the taught strategies. The results following the use of strategies 1 and 2 were consistent with the I.C.C.U.D. hypothesis.

As with Howe's results, it was difficult to determine the extent to which the information content of questions involving mole calculations affected the success rate; the results suggested that this was an important effect, although the effective redundancy of necessary information could also have made a significant contribution to the poor performance.

6.3 Garforth's Investigation of Difficulties in Using and Understanding Simple Ionic Equations

Garforth's study was motivated by the apparent inability of pupils in their pre- 'O' level year to write or interpret ionic equations. Two sets of results from her study will be considered in this Section. The first set records the preferences of English pupils in their final year of an 'O' level Chemistry course for different types of equations. As the types of equation vary in their information content, these results are of interest in considering the I.C.C.U.D. hypothesis.

Secondly, some results will be considered from a test designed to examine pupils' (aged 15+, 16+, 17+) understanding of certain principles and concepts involved in using and writing simple ionic equations, and to determine which concepts were making a major contribution to pupils' difficulties. These results, taken in conjunction with the preferences indicated for different equation types, were examined to determine pupils' apparent levels of conceptual understanding.

6.31 Pupils' Preference for Different Types of Equation

To determine 'O' level pupils' preferences for different types of equation, Garforth's test listed a series of examples, involving seven types of reaction. In each case, pupils were asked to indicate which of

TABLE 6.3

PREFERENCES FOR EQUATION TYPES

Reaction	Percentage of Pupils Choosing			
	Formal	Full Ionic	Nett Ionic	Half Equation
Carbonate/Acid	83.3	9.1	7.3	-
Neutralisation	81.6	10.0	8.1	-
Metal/Acid	79.3	9.9	6.6	4.1
Precipitation	73.0	9.6	17.1	-
Metal Displacement	69.0	11.0	15.4	4.6
Redox	65.9	9.5	20.2	6.9
Ammonium salt/ Alkali	66.1	10.0	24.3	-

the equations given best described the reaction, in their opinion. Four types of equation - each correct - were given for each example, e.g.

Zinc + Copper (II) (Cupric) sulphate solution:

A. $Zn + Cu^{2+} \rightarrow Zn^{2+} + Cu$ (Nett ionic)

B. $Zn - 2e^{-} \rightarrow Zn^{2+}$ $(Zn \rightarrow Zn^{2+} + 2e^{-})$

$Cu^{2+} + 2e^{-} \rightarrow Cu$ (Half-equations)

C. $Zn + Cu^{2+} + SO_4^{2-} \rightarrow Zn^{2+} + Cu + SO_4^{2-}$ (Full ionic)

D. $Zn + CuSO_4 \rightarrow ZnSO_4 + Cu$ (Formal)

The percentage preferences Garforth recorded for the different reaction types are shown in Table 6.3. Of course, these results may not reflect pure "personal bias", as some types of reaction, or some particular reactions, may be frequently represented by a particular type of equation. (For example, 31% recorded a preference for the nett ionic equation for $KMnO_4$ + $FeSO_4$, which was double the figure for the other redox reactions). Nevertheless, there was an overwhelming preference shown for the formal type of equation in all cases.

The fact that formal equations are preferred to full ionic equations is consistent with the idea that increasing the amount of information in a block decreases the "understandability" of that block.

The choice for full ionic equations seemed very constant for all types of reaction - in other words, different types of reaction called for a choice between formal and nett ionic (or half-equations). The choice between these alternatives is particularly interesting. The nett ionic (or half-equation) expresses concisely and exactly what happens in a given ionic reaction - that is, it gives the minimum information description of the reaction. Yet pupils preferred a type of equation which contained more information, and in which the additional information was logically redundant in that it was superfluous. The superfluity of the additional information is best shown by an example. Neutralisation is described by $H^+ + OH^- \rightarrow H_2O$; a specification of the particular spectator ions involved is surely redundant information.

A preference for formal rather than full ionic equations would be consistent with two possibilities:

(a) Some of the information in the full ionic equation being effectively redundant. The full ionic equation would then seem an unnecessarily complicated version of the formal equation; or

(b) Chunk capacity having been exceeded, the full ionic equation would be largely meaningless.

Both these possibilities are consistent with the results of Duncan and Howe.

A preference for formal rather than nett ionic (or half-equations) suggests a rather different possibility. It may be that the formal equation represents the pupils' view of "chemical reaction", while the nett ionic does not. In a way, the formal equation describes what a pupil does and sees during the course of a reaction, rather than the reaction itself. For example, if a pupil puts a piece of copper wire into silver nitrate solution, he can observe the appearance of silver crystals and the characteristic colour of copper (nitrate) solution. $Cu + AgNO_3 \rightarrow Ag + Cu(NO_3)_2$ is then a meaningful description of the process he has observed. Although he may abstract sufficiently to see this as an example of metal displacement (a concrete description of what he has observed), it does not follow that he appreciates that $Cu + Ag^+ \rightarrow Cu^{2+} + Ag$ describes the reaction underlying his observations. This apparent concern with the reactants and products, rather than the reaction itself, seems similar to the pupils' treatment of organic equations noted during the interview (Section 5.23).

We would suggest that the results quoted indicate that for many of the pupils involved in this study, the concept of ionic chemical reaction is at the rather concrete level described by the formal equation, rather than that assumed by the nett ionic equation (or half-equation). In other words, pupils have shown a lack of conceptual understanding of ionic reactions by choosing to use redundant information. (The above argument would also apply if full ionic equations were preferred to nett ionic equations. However, in this case, the choice would be consistent with a higher level of understanding of ionic reaction, but with a preference for the explicit statement of the non-involvement of the spectator ions).

Evidence supporting the view outlined above was found in the second set of results to be discussed here. Garforth suggested that thirteen concepts were involved in the understanding of ionic equations and their use. In this test, pupils (534 aged 15+, 292 aged 16+ and 182 aged 17+) were given information about seven of these concepts, namely the valencies of elements, their symbols, whether they were metals or nonmetals, a definition of the electron, a list of ionic and covalent substances, and a list of soluble and insoluble substances. Pupils were asked to choose from a given set

(a) the equation representing a given reaction

or (b) the simplest equation representing a reaction

or (c) the substances to be used to bring about the reaction specified by a nett ionic equation.

According to the distractor chosen lack of understanding (or recall) of one or more of the other six concepts - size and sign of the charge on an ion, electrical neutrality of elements and compounds, ionic equations must balance in mass and charge, and the role of spectator ions - was inferred.

In the present context, we are particularly interested in the series of questions relating to the reactions listed in Table 6.3, which also contained a spectator ion distractor, e.g.

Which of the following ionic equations represents the reaction between Aluminium sulphate solution and Barium chloride solution?

A. $Ba^{2+} + 2SO_4^{2-} \rightarrow Ba(SO_4)_2$ (Imbalance in charge)

B. $Ba^{+} + SO_4^{2-} \rightarrow Ba_2(SO_4)$ (Imbalance in mass, wrong charge)

C. $Al^{3+} + 3Cl^{-} \rightarrow AlCl_3$ (Spectator ion)

D. $Ba^{2+} + SO_4^{2-} \rightarrow BaSO_4$ (Key)

E. $2Ba^{+} + SO_4^{2-} \rightarrow Ba_2SO_4$ (Wrong charge)

The results in Table 6.4 indicate that on average, over 60% of pupils chose either the key or the spectator ion distractor for each type of reaction. Given the understanding of chemical reaction suggested above, neither the key nor the spectator ion alone would describe the reaction satisfactorily; the choice between the two should depend on their relative importance (from the pupils' point of view) in different types of reaction. Reactions in which spectator ions should be less important are

(i) precipitation - in which the precipitate would seem most important;

(ii) ammonia salt/alkali - from the pattern "ammonia salt + alkali → ammonia";

(iii) carbonate/acid, again from the pattern "carbonate + acid → carbon dioxide".

TABLE 6.4

CHOICE OF CORRECT RESPONSE AND SPECTATOR ION DISTRACTOR

Reaction / Age	% Correct			% Choosing Spectator Ion Distractor		
	15+	16+	17+	15+	16+	17+
Precipitation	44.7	64.8	64.5	25.6	16.6	14.7
Ammonium salt/OH^-	34.4	59.9	73.0	29.7	14.1	9.5
Carbonate/Acid	25.3	49.9	59.8	36.4	22.0	18.5
Neutralisation	22.9	44.6	57.5	52.1	38.6	26.5
Redox	18.2	40.6	53.0	46.1	27.0	23.0
Metal/Acid	13.6	35.7	54.0	* 19.4	19.3	11.8
				** 51.0	32.0	23.2

* Distractors also contained a non-neutral element.

** Distractors also contained an imbalance in charge.

Reactions in which spectator ions could be more important are

(i) neutralisation - from the pattern "acid + base → salt + water"; i.e. a confusion between neutralisation and salt formation;

(ii) metal + acid.

In the latter instances, pupils would be choosing in favour of the product specific to a given example of a reaction, rather than the product common to all examples of the reaction. With the exception of redox reactions (which cannot readily be fitted into either of these categories) the relative choice between spectator ion and key followed the pattern outlined above. It is very interesting to note that a marked shift in the proportion choosing the spectator ion distractor to the key response occurred in all cases except precipitation reactions, with increase in age. The pattern of these results is consistent with the level of conceptual understanding outlined above; it also suggested that an increase in conceptual understanding had occurred in the 17+ age group.

According to the I.C.C.U.D. hypothesis, if a pupil with this rather concrete understanding of chemical reaction studied a full ionic or nett ionic equation, he would be doubly disadvantaged:

(a) Lack of conceptual understanding of the components of the equation could inhibit an holistic interpretation of the equation, or the use of a convergent strategy;

(b) Such information as he did extract could, given the understanding of chemical reaction outlined above, be effectively redundant, which would further inhibit meaningful interpretation of the equation. The preference for formal equations can, in the light of the understanding shown, be described as a preference for the minimum (as judged by pupils) information expression of a situation. This is an important result, which is consistent with the I.C.C.U.D. hypothesis.

6.4 Conclusions

The examination of each of these sets of results provided a fairly clear picture of pupils' levels of conceptual understanding.

From Howe's results, it was possible to calculate Redundancy figures within many of the stages in his structured test. These figures indicated that a considerable percentage of pupils, when required to write simple binary formulae, found basic chemical information effectively redundant. This certainly seemed to suggest a low level of conceptual understanding of electron configuration, and bonding, and thus of ions and the formation and composition of simple compounds. A further redundancy calculation indicated a low level of conceptual understanding of the mole.

Duncan's results also seemed to give a clear indication of a low level of conceptual understanding of the mole, and, in addition, of molarity and concentration. Indeed, in the relevant sections of Duncan's test, high F.V.'s were observed only in situations in which a simple mathematical strategy could be used.

Garforth's test was structured very differently from the other two, and we considered a more specific set of results, but here too the results suggested strongly that there was a low level of conceptual understanding of chemical reaction. This was indicated by the high percentage of pupils choosing a formal equation descriptor for the different reaction types presented; an indication that was reinforced by the observed pattern of selection of spectator ion distractors for the different reactions.

It is also important to notê that each of these authors concluded that their complete results indicated that very many pupils had failed to attain key concepts within their area of study. Given this lack of relevant conceptual understanding, what of pupils' performance in varying information contexts?

Howe's and Duncan's results provided a variety of situations in which comparisons of success rate could be made as the information content or memory load increased. Two numerical comparisons were possible. Firstly, Duncan's results gave the F.V.'s for the individual steps of a weight-for-weight calculation; these were used to compute an "expected" F.V. for the combined computation, which clearly exceeded the recorded F.V. Secondly, Duncan's results showed that pupils taught Strategy 1 and those taught Strategy 2 performed poorly in the relevant post-test, as we had predicted, but that Strategy 2 - which imposed the higher memory load at a crucial stage - was distinctly less helpful than Strategy 1. In both Howe's and Duncan's results, there were several examples of pupils' achieving reasonable success rates in low information situations, or situations in which a simple (usually mathematical) strategy could be employed, but showing much poorer performances in similar, higher information contexts, or contexts in which their strategies could not be used. Howe's results also included pupils' difficulty ratings for the tasks required of them in his tests, and generally, the higher information content tasks drew the higher difficulty ratings. We also found that higher information tasks were associated with a greater decrement in performance going from an immediate to a delayed test.

There was an indication that, in some situations, effective redundancy of information could have made an additional contribution to the poor performances noted. One such situation involved computations using equations that did not have 1:1 mole ratios. Now, following the argument in Section 6.21, taught strategies for n:m mole ratio problems must impose a heavier memory load simply because an additional computational step is required, and so they represent higher information contexts, even though the amount of given information is the same as in 1:1 ratio problems. It may also be that pupils choose to recall only the simple 1:1 ratio rule, and to ignore the given equation - a strategy that would enable them to be successful at least some of the time. These two possibilities represent one direct, and one indirect, reaction to the information content of the situation. However, while it would be important to determine whether (or in what circumstances) effective redundancy makes an entirely independent contribution to the difficulty of a task, such a distinction is not relevant in the present case; we would make no claim that information content is the only factor that affects performance, given low conceptual understanding.

The information dependent section of Garforth's results considered related to just one type of performance - pupils' selection of a description, given several possible equations that differed in information content. It was our view that this was a very important aspect of learners' behaviour, which is seldom studied. Her results seemed to indicate clearly that pupils' preference was for the equation type that gave the minimum information description (from their point of view) of the situation.

In all the cases examined here, only two counter examples were noted - and one of these (Section 6.223) related to the simple number effect, while the other (6.222) could have involved a mathematical strategy difference. One of the most important findings of this examination is that the overall, general trend is that of apparent consistency with the predictions of the I.C.C.U.D. hypothesis. It was our view that if the I.C.C.U.D. hypothesis did not appear to have general relevance, its use in explaining a specific set of results (such as the Organic Chemistry results) would be very questionable.

In summary, then, the examination of these results found substantial consistency with the predictions of the I.C.C.U.D. hypothesis. The implications of this finding will be considered in the next chapter.

APPENDIX 6.1

Howe's Test Items and Results

ITEM NO.	SUMMARY OF QUESTION	I % CORRECT	II % ALSO CORRECT IN PRIOR STEPS	III % REDUNDANCY	
3,4	Prior step: (3) Chose correct electron arrangement for $^{16}_{8}O$.	73			
	Final step: (4) Atoms seem more stable when they have a completely filled outside level (shell). The $^{16}_{8}O$ atom might achieve this by				
	(a) ... two electrons	a. 62	3. 49	(3-4a)	36
	or (b) ... two electrons.	b. 71	3. 53	(3-4b)	37
	(a) = sharing, (b) = gaining.	a+b.43	33	(3-{a+b})	50
5,6	Given four atomic symbols (e.g. $_{9}F$)				
	Prior steps: (5a) State electron arrangement (3/4)	76			
	(5b) State valence No. (3/4)	66	57	(5a-5b)	29
				(5a-6)	?
				(5b-6)	?
	Final Step: (6) Write the formulae of four specified compounds containing these atoms. (3/4)	43	35		
	More than one wrong, but correct method.	20	?		
7,8	Given names of 4 compounds (e.g. Silicon Hydride)				
	Prior step: (7) Name the elements in each compound (3/4)	70			
	Final step: (8) Use the periodic table and write the formulae of these compounds. (3/4)	33	31	(7-8)	42
	More than one wrong, but correct method	19	13		

ITEM NO.	SUMMARY OF QUESTION	I % CORRECT	II % ALSO CORRECT IN PRIOR STEPS	III % REDUNDANCY
9,10	Given: The formula of the sodium ion, and the nett charge on a compound is zero,			
	Prior steps: Complete $_{17}Cl$ + (9a).. $_{17}Cl^{(9b)..}_{2.88}$	a. 69 b. 48	 41	(9a-10) 31 (9a-9b) 35
	Final step: (10) Write the formula for Sodium Chloride.	80	36	(9b-10) 46
11,14	Prior steps: (11) Calcium is in column 2 of the periodic table. Write the electron arrangement for $_{20}Ca$.	78		
	(12) How many electrons will it lose to have a completely filled outside shell.	81	73	(11-12) 13 (11-13) 43
	(13) Write the symbol for the Calcium ion.	51	42	(12-13) 41 (12-14) 30
	Final step: (14) Write the formula for Calcium Chloride.	69	36	(13-14) 36
15	Given: Name and charge on 6 ions.			
	Write formulae (name + charge) of four compounds. (3/4)	63		
16,17	Prior steps: (16) Formulae (symbols and charge) of five oxyanions. (3/5)	50		
	Final steps: (17a-d) Formulae for four compounds containing anions. (3/4)	27	24	
	(17e) Formula for Sodium Silicate.	4	(17a-d) 3	(17(a-d)-17e)25

ITEM NO.	SUMMARY OF QUESTION	I % CORRECT	II % ALSO CORRECT IN PRIOR STEPS	III % REDUNDANCY
18,19	Prior step: (19) How many moles (gram atoms) of iron are there in 1 mole of Fe_2O_3.	32		
	Final step: (18) Calculate the formula weights of four compounds. (3/4 or method correct)	81	29	(19-18) 58
20	Translate into words: $Na + S \rightarrow Na_2S$	30		
21	What is meant by a 'balanced equation'?	30		
22	When is it essential to write a balanced equation?	23		
23	Given $2Pb(NO_3)_2 \rightarrow 2PbO + 4NO_2 + O_2$			
	(a) How many moles of NO_2 from 2 moles of $Pb(NO_3)_2$?	20		
	(b) How many moles of O_2 from 1 mole of $Pb(NO_3)_2$?	3		
	(a) and (b) correct	1.5		
24	Given $CaCO_3 \rightarrow CaO + CO_2$ F.Wt. 100 56 44 What weight of CO_2 would be made by completely roasting 15 g. chalk? Prior step: 23(b)	30	2	(23b-24) 29

ITEM NO.	SUMMARY OF QUESTION	I % CORRECT	II % ALSO CORRECT IN PRIOR STEPS	III % REDUNDANCY
25	Balance:			
	(a) $Ca(OH)_2 + HCl \rightarrow CaCl_2 + H_2O$	27		
	(b) $Ca^{2+}_{(aq)} + 2Cl^-_{(aq)} + Ag^+_{(aq)} + NO^-_{3(aq)} \rightarrow Ca^{2+}_{(aq)} + 2NO^-_{3(aq)} + Ag^+Cl^-_{(s)}$	32		
	both (a) and (b) correct	18		
26	Rewrite the sentences using only words			
	(a) He sprinkled NaCl into a beaker full of H_2O.	81		
	(b) $H_2O_{(l)}$ and $Mg_{(s)}$ do not react readily to give $H_{2(g)}$ but when $H_2O_{(g)}$ is passed over hot $Mg_{(s)}$ it gives $H_{2(g)}$ and $MgO_{(s)}$.	33		
	(c) $Ag^+_{(aq)} + NO^-_{3(aq)}$ were added to $H^+_{(aq)} + Cl^-_{(aq)}$ to give $Ag^+Cl^-_{(s)} + H^+_{(aq)} + NO^-_{3(aq)}$.	5		
	(a) and (b) correct	31		

APPENDIX 6.2

Duncan's Test Items and Results

The Facility and Discriminating Power are given for each item used. The D.P. is shown in brackets below the F.V.

	Items	Programme Group	Class Group
1.	A molar solution of HCl contains	0.28 (0.23)	0.58 (0.37)
2.	Which of these HCl solutions is most concentrated? (e.g. "500 ml of 2M HCl")	0.44 (0.36)	0.50 (0.44)
3.	Which solution of NaCl is most concentrated? (e.g. "200 ml of solution containing 2 moles of NaCl")	0.49 (0.20)	0.57 (0.32)
4.	If one mole of Sodium Hydroxide (NaOH) is dissolved in 500 ml of solution, what is the concentration?	0.38 (0.48)	0.56 (0.25)
5.	If .5 moles of NaOH are dissolved in 200 ml of solution, what is the concentration of the solution?	0.56 (0.57)	0.77 (0.31)
6.	Which solution contains most NaCl? (e.g. "500 ml of 2M NaCl")	0.48 (0.14)	0.51 (0.37)
7.	Which of the following solutions contains most NaCl? (e.g. "30 ml of 1.2M NaCl")	0.34 (0.24)	0.35 (0.30)
8.	How many moles of NaOH are dissolved in 500 ml of 4M NaOH?	0.64 (0.57)	0.81 (0.33)
9.	How many moles of H_2SO_4 are dissolved in 15 ml of 2M H_2SO_4?	0.32 (0.41)	0.58 (0.30)
10.	What weight of NaOH is contained in 500 ml (½l) of 1M NaOH?	0.61 (0.40)	0.76 (0.42)
11.	What weight of NaOH is contained in 100 ml of 5M NaOH?	0.56 (0.41)	0.72 (0.43)
12.	How many moles of NaOH react with 1 mole of H_2SO_4? (Given: $2NaOH + H_2SO_4 \rightarrow Na_2SO_4 + 2H_2O$)	0.58 (0.41)	0.76 (0.19)

	Items	Programme Group	Class Group
13.	How many moles of NO_2 from 1 mole of $Pb(NO_3)_2$? (Given: $2Pb(NO_3)_2 \rightarrow 2PbO + 4NO_2 + O_2$)	0.46 (0.12)	0.50 (0.37)
14.	How many moles of H_2 react with 1 mole of N_2? (Given: $N_2 + H_2 \rightarrow NH_3$)	0.41 (0.44)	0.54 (0.64)
15.	What weight of Mg reacts with 32g S? (Given: $Mg + S \rightarrow MgS$)	0.66 (0.57)	0.74 (0.45)
16.	What weight of SO_2 reacts with 32g O_2? (Given: $2SO_2 + O_2 \rightarrow 2SO_3$)	0.24 (0.29)	0.34 (0.52)
17.	What weight of O_2 reacts with 3g C? (Given: $C + O_2 \rightarrow CO_2$)	0.65 (0.43)	0.77 (0.44)
18.	What weight of Al reacts completely with 80g CuO? (Given: $2Al + 3CuO \rightarrow Al_2O_3 + 3Cu$)	0.27 (0.30)	0.30 (0.43)
	Given: $Mg + H_2SO_4 \rightarrow MgSO_4 + H_2$		
19.	How many moles of Mg react with 1l. 1M H_2SO_4?	0.89 (0.21)	0.90 (0.31)
20.	How many moles of Mg react with 100 ml 4M H_2SO_4?	0.40 (0.42)	0.78 (0.35)
21.	What volume of 2M H_2SO_4 reacts with 2 moles of Mg?	0.40 (0.37)	0.64 (0.35)
22.	What volume of 4M H_2SO_4 reacts with ½ mole of Mg?	0.39 (0.40)	0.62 (0.46)
	Given: $NaOH + HCl \rightarrow NaCl + H_2O$		
23.	What volume of 1M NaOH reacts with 2l. of 1M HCl?	0.69 (0.43)	0.81 (0.34)
24.	What volume of 4M HCl reacts with 80 ml of 1M NaOH?	0.66 (0.44)	0.79 (0.31)
25a.	½l of 1M NaOH is neutralised by 1l. of HCl solution. What is its molarity?	0.50 (0.28)	0.61 (0.41)
25b.	25 ml of 4M HCl is neutralised by 100 ml of NaOH solution? What is its molarity?	0.43 (0.51)	0.74 (0.33)

Items	Programme Group	Class Group
Given: $2NaOH + H_2SO_4 \rightarrow Na_2SO_4 + 2H_2O$		
26. What volume of 1M NaOH will neutralise 1l. of 1M H_2SO_4?	0.36 (0.30)	0.58 (0.38)
27. What volume of 2M H_2SO_4 neutralises 250 ml of 4M NaOH?	0.24 (0.08)	0.23 (0.38)
28. 1l. of 1M NaOH neutralises ½l. of H_2SO_4. What is the molarity of the H_2SO_4?	0.22 (0.23)	0.27 (0.27)
29. 20 ml of 2M H_2SO_4 neutralise 100 ml NaOH. What is the molarity of NaOH?	0.16 (0.19)	0.19 (0.37)

CHAPTER 7

General Conclusions, Implications and Recommendations

7.1 A Review

Following discussions with teachers and pupils, two hypotheses were proposed to account for the difficulties pupils had reported in the learning of Condensation, Hydrolysis and Esterification Reactions (C.H.E. Reactions). The first of these, the Visual Difficulties Hypothesis, encapsulated the teachers' view that pupils' difficulties in this area arose because they were in some way confused by the pattern characteristics of formulae - in particular, by the sort of "back-to-front" representations of formulae used in C.H.E. equations. The second hypothesis proposed that pupils' difficulties arose because of a lack of conceptual understanding of functional groups. The recognition as units of the common functional groups and the sub-units $-CH_3$ and $>CH_2$ (termed the "Class I" groups), and the noting of the functional group of a formula first, were proposed as observable behaviours indicative of acceptable and useful levels of conceptual understanding of functional groups.

The Combined Tests Experiment was designed as a critical test of these two hypotheses. The results of this experiment contra-indicated the Visual Difficulties Hypothesis. Pupils' ability to process and recall formula-pattern information seemed to correspond to Short Term Memory capacity for other types of information, and their incorrect responses were incomplete or inaccurate, but not confused, representations, of the original Test Patterns. There was one exception to this - the "box" pattern shown by certain formulae (Section 4.41). We have already recommended that these formulae be written in such a way that this pattern characteristic is suppressed. This experiment produced the further interesting result that pupils tended to "read" patterns and formulae from left to right, as though they were two-dimensional words.

The Combined Tests results also indicated that very few pupils exhibited the criterial recognition behaviour. Even at SYS level, fewer than 10% of the sample recognised as units the Class I groups, and a substantial proportion of the fourth year pupils (15%) recognised no groups as units. These results were established on the basis of a very stringent comparison of the Expected and Observed numbers of correct responses for the Molecule Test Items. There was, overall, no indication that pupils noted the functional group in a formula first (although the small number of Class I pupils rendered this conclusion less certain in their particular case).

Because of the very small number of Class I pupils, only limited, low power tests of the correlation between recognition (as measured by the Class I ratio) and performance in Chemistry could be made. The results of these tests were consistent with the proposition that the criterial behaviour was a valid indicator of conceptual understanding, but they could not be classed as strong evidence of this validity because of their low power.

The results of the Grid Test Experiment enabled a more detailed investigation to be made of pupils' recognition behaviour and their performance in common Organic Chemistry tasks. These results indicated that in this experimental situation also, pupils did not treat functional groups as units. They further demonstrated a failure to choose the functional group as a characteristic property and suggested that pupils were not strongly committed to the view that chemical behaviour was specifically relatable to functional groups.

These results certainly suggested that many pupils had not attained a level of conceptual understanding in which a functional group was considered as a unit - and an important behaviour determining unit. Furthermore, we noted that the mistakes pupils made followed consistent patterns, directly relatable to this low level of conceptual understanding, and that the effect of this lack of conceptual understanding was reduced by the use of specially designed learning materials. The results of a small scale interview supported the interpretation of the Grid Test results, and provided some interesting examples of strategies used in interpreting organic equations.

Taken together, the two sets of experimental results provided strong support for the hypothesis that pupils' difficulties were conceptual in origin.

The I.C.C.U.D. hypothesis was proposed as a model that would explain why pupils judged C.H.E. reactions to be a difficult topic, when their lack of conceptual understanding would seem to be relevant in many areas of Organic Chemistry. An examination of the results from three independent studies of areas of difficulty in Chemistry showed substantial consistency with the predictions of the I.C.C.U.D. hypothesis. This would suggest that pupils' rating of C.H.E. reactions was in fact significantly related to the high information content of the tasks characteristic of this topic.

And so it would seem that teachers were quite correct in suspecting that the formulae and equations involved in C.H.E. reactions were related to pupils' difficulties - but it would appear that it is their information content (given a low level of conceptual understanding) and not their pattern characteristics that is important.

At this stage, we would wish to make some general proposals related to the I.C.C.U.D. hypothesis, and some specific recommendations concerning the teaching and learning of Organic Chemistry. These will be presented separately in the following sections.

7.2 The Implications of the I.C.C.U.D. Hypothesis

The experimental results analysed in Chapter 6 revealed a degree of consistency with the predictions of the I.C.C.U.D. hypothesis which, we believe, would warrant the mounting of a study specifically designed as a critical test of the hypothesis.

Howe's and Duncan's results indicated the suitability of problem solving tasks for such a study, and they also indicated certain factors (such as the simple number effect, and the use of mathematical strategies) that should be considered carefully in the design phase of a critical test. In particular, it would be necessary to ensure that any effect due to the effective redundancy of necessary information was not confounded with an effect due to change in information content. It would also seem that the measurement of information redundancy within a logically related sequence of steps could provide a useful tool for determining levels of conceptual understanding - the required first phase of a critical experiment.

We would suggest that the I.C.C.U.D. hypothesis merits critical testing because it has potentially important implications for the understanding of learning processes, and for teaching practice. We have already suggested (in Section 5.32) that this hypothesis implies a vicious circle in the learning of Organic Chemistry. This implication would be quite general, and so the same sort of vicious circle would be expected in other areas of Chemistry, where conceptual understanding must be built up at least in part from learning in high information contexts.

The I.C.C.U.D. hypothesis proposes a fairly detailed description of the way in which a learner's level(s) of relevant conceptual understanding, by mediating his interaction with the information content of a new learning task, influence the development of his conceptual understanding. In this respect, we see a close and interesting parallel between this model and Ausubel's theory of meaningful learning. Ausubel considers that meaningful learning can occur when the learner possesses subsumers (this type of learning has been considered briefly in Section 2.4), and that these subsumers function by structuring or organising new material so that it can be readily and meaningfully absorbed into existing cognitive structure. He further states that the subsumers themselves may acquire new generic meaning following meaningful learning.[91] It would seem that the I.C.C.U.D. hypothesis describes one quite direct way in which a learner's existing level(s) of conceptual understanding (his subsumers) structure new learning material, and thereby affect the possibility of meaningful learning's occurring.

Because of the detailed relations it proposes, the I.C.C.U.D. hypothesis also indicates ways of alleviating the learning difficulties that may be expected to occur during the period in which at least a minimal level of conceptual understanding is being acquired. According to the I.C.C.U.D. hypothesis, pupils with a low level of conceptual understanding are disadvantaged because:

(a) they chunk inefficiently; that is, they form chunks of low information content;

(b) they may increase memory load by treating redundant information as necessary;

(c) they are liable to use inefficient or arbitrary strategies in high information contexts.

These disadvantages could be lessened by:

(i) increasing chunking efficiency (which would effectively decrease the number of chunks required to represent a given amount of information);

(ii) providing pupils with simple, efficient strategies for use in high information contexts (which could be computational rules, or strategies that enabled pupils to identify salient information);

(iii) reducing the total amount of information that must be considered.

The results discussed in Chapter 6 provide some interesting points that illustrate and clarify the second and third proposals. (An application of the first proposal will be considered in the next Section). A reduction in the total amount of information to be considered could be achieved by simply excluding - or delaying - the teaching of certain topics. For example, if volumetric calculations are included in a syllabus primarily to enhance conceptual understanding of the mole and perhaps molarity, the I.C.C.U.D. hypothesis would suggest that they cannot fulfil this function in a population that is shown to have a low conceptual understanding of the mole. The discussion of Section 6.21 strongly suggested that, given low conceptual understanding, solution rules such as Strategy 1 and Strategy 2 would impose a severe load on Working Memory; a load which could be increased still further during the performance of computational steps. In circumstances such as this, it would seem in the best interests of pupils to delay treatment of such a topic until some conceptual understanding had been acquired through the use of lower information contexts.

In the general case, where a topic cannot be delayed or excluded, a simple reduction of information - or a minimisation of total information to be considered - will not necessarily be successful. Garforth's results provided a very nice illustration of this point. Her results showed that pupils did not choose the minimum total information description (the nett ionic equation) but rather the minimum information description that they deemed to be the least arbitrary (the formal equation). Now, if we were to minimize only the arbitrary information - by considering only formal equations - reinforcement of pupil's misunderstanding of ionic reactions could easily result. If however the total information were minimized - by using only nett ionic equations - the pupils might judge the information given to be incomplete or arbitrary, which could again inhibit their acquisition of conceptual understanding.

Bearing in mind that the limitation on information storage is a constraint on the number of chunks, the I.C.C.U.D. hypothesis would suggest that the information to be considered should be reduced specifically by conjointly minimising the total information and the arbitrary information given. The judgement of "arbitrary" (or effectively redundant) must be made from the pupils' demonstrated point of view; the "minimum" must be the minimum necessary to achieve the educator's purpose. Such a reduction process may still require the consideration of a reasonable amount of information, and in these cases, the use of a taught strategy would be advantageous.

It would seem that the most successful taught strategies for use during the period of acquisition of conceptual understanding would be those that conform to the above minimisation specification. The apparent redundancy of electron configuration in the writing of binary formulae, shown by Howe's results, provides an interesting illustration of this point. It would seem that students could use this information to determine the formula of a compound only if they were to use a taught strategy. Such a taught strategy would necessarily involve the specification of quite a few steps (for example, "deduce the inert configuration closest to the given configuration"), and so would be a relatively high information rule. Howe's results suggested quite specifically that some of the given information would be effectively redundant for some of these steps - for example, the deduction of ionic charge. In these circumstances, it would seem that pupils would be more successful in writing formulae if they were initially taught a simpler rule, which conformed better to the minimisation criteria outlined above, and considered the relation between electronic structure and bonding at a later stage.

It is very interesting to note that both Howe and Duncan recommended that volumetric calculations be deferred until year 5 of the 'H' Grade course and also made proposals that effectively required the information content of problem areas to be reduced. It is also the case that Memorandum Paper No.3[(11)] recommended the use of simple formulae and equations wherever possible (and thus effectively suggested that their information content be reduced).

Thus the I.C.C.U.D. hypothesis, in describing a particular relationship between information content, conceptual understanding and difficulty provides a model which may increase our understanding of the causes of pupils' learning difficulties (in certain areas) and also suggests specific guidelines for information manipulation and minimisation that could lessen these learning difficulties in high information situations. It is for these reasons that we recommend a stringent test of the hypothesis.

7.3 Recommendations for the Teaching of Organic Chemistry

In Section 2.4 we suggested that the learning of science concepts is very properly regarded as a desirable end in itself, as well as being of potential use in the learning of new material. For this reason, the recommendations that we shall make in this section have the joint purpose of lessening the apparent learning difficulties within the topic of C.H.E. reactions, and of improving the level of conceptual understanding of functional groups - surely a key concept in Organic Chemistry.

Even at SYS level few pupils recognised Class I groups as units, and so it would seem that this conceptual problem is not a matter of maturity - or else it is a very severe maturity problem, not resolved till post-secondary stages. Johnstone's original survey[(9)] indicated that pupils at SYS level no longer considered the mole a difficult topic (Table 1.2). Duncan's maturity investigation (referred to in Section 6.21) showed a significant improvement in performance in his series of post-tests going from third year to fifth year pupils, which would support Johnstone's finding. While it is very difficult to compare the apparent complexity (for learners) of two different concepts, it would seem that the notion of a functional group is intrinsically both simpler and less abstract than the notion of a mole. Therefore, if pupils feel they have grasped the concept of a mole at the SYS stage, it would seem unnecessarily pessimistic to suggest that they cannot acquire a useful level of conceptual understanding of functional groups until a post-secondary stage of education.

Perhaps the problem has arisen in part because of the apparent simplicity of the notion of a functional group - to teachers. Most teachers expect that their pupils will find the concept of a mole - and related problem tasks - difficult, but as we have reported (Section 2.2) many teachers were initially surprised that pupils considered C.H.E. reactions a difficult topic. Again, during the visits to schools, teachers clearly expressed a definite opinion that this topic (and the junior Organic course generally) ought to have been suitable for their pupils. It may be that we, as teachers, have too readily assumed that the facts that a functional group is a unit, a chemical entity, whose behaviour is conserved (to a first approximation, at least), which determines family membership and chemical behaviour - all so obvious to us, are automatically equally apparent to our pupils. For example, during school visits, teachers were often observed to introduce a new organic family in the following way. First the formula of the family functional group was written on the board, and the systematic nomenclature suffix stated. Then the formulae of 5 or 6 family members would be written up, and pupils directed to write down their systematic names. This immediately drew pupils' attention away from the functional group, and could have implied that the C-H chain was the really important part of the molecule. Again, when discussing characteristic reactions, teachers commonly referred to "the properties or reactions of the ... family"; is is hardly surprising that pupils relate these reactions to families rather than functional groups. Perhaps what we must do is let our pupils into the secret, by making more explicit reference to the nature and role of functional groups.

However, the whole tenor of our discussion of the learning of science concepts (Section 2.4) argues that simply telling pupils "the facts" will not bring about instantaneous conceptual understanding. Nor will it resolve their difficulties with C.H.E. reactions. For this reason, the proposals we will make concern the use of teaching strategies or techniques that will encourage pupils to regard functional groups as units - and very important units, and will also allow them to cope with the high information content of C.H.E. equations more readily. Furthermore, the strategies and techniques suggested are intended to be of use for junior pupils, where the problem clearly begins.

Our primary proposal is concerned with effectively reducing the information content of an organic formula, and hence of organic equations. The total information of a formula could be reduced by using the conventional symbols R, R' etc., to represent the C-H chains. This would certainly highlight the functional groups of a formula. However, these symbols could well be arbitrary - or effectively redundant - from a junior pupil's point of view, and one of the implications of the I.C.C.U.D. hypothesis is that the use of such arbitrary information should be avoided. The use of this convention would also seem undesirable for junior pupils who find the connection between full structural formulae and physical molecules sufficiently tenuous without a further degree of abstraction being added. We would suggest that chains be represented by symbols only when pupils have already demonstrated that they regard them as units, whose precise details are relatively unimportant in many circumstances. At this stage, such symbols are merely the external equivalent of the learner's internal representation of a chain.

There is a simple alternative method for effectively reducing the information content of a formula - by writing the functional group in colour. This would draw pupils' attention to the functional group, implicitly suggest it was a unit, and indicate the relatively lesser importance of the details of the environment, while not removing it completely from consideration. We would suggest that this strategy conforms to the minimisation criterion derived from the I.C.C.U.D. hypothesis, in that it could lead pupils to chunk a formula more efficiently as "functional group" + "the rest", and certainly involves a minimum of arbitrary information. It is, in fact, an example of an application of the first proposal for lessening the disadvantage of low conceptual understanding given in Section 7.2.

Experimental investigation would be required to determine the effectiveness of this strategy, and also to determine whether a specific colour should be reserved for each group, or whether the use of just one contrasting colour scheme (i.e. one colour for all environments, and a contrasting colour for all functional groups) would suffice. (The latter would have clear practical advantages for both teachers and pupils). The length of time for which such a strategy should be employed would also have to be determined experimentally. This technique would, quite naturally, enable the teaching of Organic Chemistry to include a more explicit (though simple) treatment of functional groups. It would also lead naturally to the explicit teaching of an efficient strategy for interpreting organic equations, making the search for a change in functional groups (followed by a search for a change in environments) both obvious and simple. This taught strategy takes account of the implications of the I.C.C.U.D. hypothesis, in that the strategy rule is simple, and the salient information is clearly indicated. Overall, this "colour code" strategy should reduce the "information hurdle" of C.H.E. equations, and this, together with the emphasis given to the unit character of functional groups, should both facilitate the acquisition of an acceptable level of conceptual understanding of functional groups and improve pupils' performance during the acquisition period.

Earlier in this Chapter we have considered the I.C.C.U.D. hypothesis in the light of Ausubel's theory, and it is instructive to do so again. In a sense, the technique suggested above effectively pre-chunks some of the given information. This type of information manipulation could be called an organisational aid[15] in the Ausubelian sense, in that it imposes a structure on new material, increasing the likelihood of meaningful learning in spite of the lack of relevant subsumers. Information manipulation of this kind could well provide useful organisational aids in other areas of Chemistry, where the more traditional types of organisational aid, such as analogies, overviews, etc., are not relevant.

This use of colour is also similar to techniques that have been employed successfully in concept learning tasks, where a particular criterial attribute has been indicated in colour in the initial learning sessions.[20,21]

It would be very interesting to include, in a test of the "colour code" strategy, a measure of pupils' performance in identifying isomers after using the strategy. The identification of isomers often requires pupils to consider large organic formulae, and one would expect that a failure to consider functional groups as behaviour determining units could well lead to poor performance in such tasks. Thus a strategy useful in the case of C.H.E. reactions should also presumably be of use in this area.

Our second proposal concerns practical work. During the initial visits to schools, several teachers reported that they considered a series of experiments - in which pupils prepared a variety of esters - of great importance in helping pupils to understand the structure of esters. (This opinion was also noted by Gunning and Johnstone[92] in an extensive investigation of practical work in Scottish Secondary Schools). While the class discussion after the experiments considered the structure of esters, the experiments themselves certainly could not fulfil this objective, and should not be conducted for this purpose. Indeed, it was quite clear that during the experiments pupils were concerned with only one property of esters - their smell. On the other hand, the use of molecular models kits by pupils did seem to be very effective in encouraging them to consider the structure of molecules, and in demonstrating in a very concrete way, the involvement of functional groups in chemical reaction. Teachers certainly favoured their use, but, because of the cost of these kits, were generally unable to provide a class set for regular use. We would recommend that high priority be given to the acquisition and use of such kits.

The card game, developed for use in the Grid Test Experiment, was shown to be effective in encouraging pupils to choose the functional group as a characteristic property, and to treat it as a unit. (The experimental version was developed for fifth year pupils, but the packs can be readily adapted for fourth year use by replacing the primary, secondary and tertiary alcohol families with a single 'alcohol' family). This card game could provide a further useful learning exercise for pupils. (Mr Young of Dalziel High School, Motherwell, who participated in the experimental phases of this study, has made the valuable suggestion that the card game would be more useful if the family name cards showed the systematic nomenclature suffix associated with the family).

While the results of this study may well have relevance to many areas of difficulty in Chemistry, its specific finding has been that at present, few pupils have achieved a useful level of conceptual understanding of functional groups, and that this has contributed in a major way to learning difficulties in one area of Organic Chemistry. It is hoped that the results of this study will prove useful in remedying this situation, allowing pupils to appreciate more fully this very interesting and important area of Chemistry. Perhaps it has also shown that if we wish pupils to acquire our view of Chemistry, we must first be prepared to look at it through their eyes.

REFERENCES

1. R.W. Tyler, J. Res. Sci. Teach., 1967, 5, 52.

2. Alternative Chemistry Syllabuses Ordinary and Higher Grades, Scottish Education Department - Circular 512, October 1962.

3. B. Inhelder and J. Piaget, The growth of logical thinking, London: Routledge and Kegan Paul, 1958.

4. R.B. Ingle and M. Shayer, Educ. Chem., 1971, 8, 182.

5. M. Shayer, Educ. Chem., 1970, 7, 182.

6. K. Lovell, Studies in Sci. Ed., 1974, 1, 1.

7. L.G. Dale, Aust. J. Psych., 1970, 22, 277.

8. A.H. Johnstone, Studies in Sci. Ed., 1974, 1, 21.

9. A.H. Johnstone, T.I. Morrison and D.W.A. Sharp, Educ. Chem., 1971, 8, 212.

10. G.V. Glass and J.C. Stanley, Statistical methods in education and psychology, New Jersey: Prentice-Hall, 1970, Chapt. 14.

11. Memorandum No. 3, National Curriculum Centre for Mathematics and Science, Dundee, 1971.

12. P.J. Fensham, "Concept formation", in New movements in the study and teaching of chemistry, D.J. Daniels (Ed.), London: Temple-Smith, 1976.

13. J.D. Novak, D.G. Ring and P. Tamir, Sci. Educ., 1971, 55, 483.

14. D.P. Ausubel, Educational psychology - a cognitive view, Holt, Rinehart and Winston, 1968, p. 24.

15. L.H.T. West and P.J. Fensham, Studies in Sci. Educ., 1974, 1, 61.

16. L.H.T. West, J. Res. Sci. Teach., 1976, 13, 297.

17. W.E. Vinacke, Psych. Bull., 1951, 48, 1.

18. E. Heidbreder, J. Psych., 1949, 27, 263.

19. B.E. Shepp and D. Zeaman, J. Comp. and Phys. Psych., 1966, 62, 55.

20. C.L. Hull, Psych. Monographs, 28, Whole Number 123.

21. T.R. Trabasso, J. Exp. Psych., 1963, 65, 398.

22. L.E. Bourne Jr and R.C. Haygood, J. Exp. Psych., 1959, 58, 232.

23. L.E. Bourne Jr and R.C. Haygood, J. Exp. Psych., 1961, 61, 259.

24. R.N. Sheppard, G.I. Hovland and H.M. Jenkins, Psych. Monographs, 1961, 75, No. 13.

25. T.R. Trabasso and G.H. Bower, Attention in learning: theory and research, New York: John Wiley, 1968.

26. W.F. Battig and L.E. Bourne Jr, J. Exp. Psych., 1961, 61, 329.

27. J.S. Bruner, J.J. Goodnow and G. Austin, A study of thinking, New York: John Wiley and Sons Inc., 1967.

28. R.C. Haygood and L.E. Bourne Jr., Psych. Rev., 1965, 72, 175.

29. L.E. Bourne Jr, "Learning and utilization of conceptual rules" in Concepts and the structure of memory, B. Kleinmuntz (Ed.), New York: Wiley, 1967.

30. L.W. Gregg, "Internal representations of sequential concepts", ibid.

31. T.R. Trabasso and G.H. Bower, J. Math. Psych., 1966, 3, 163.

32. F. Restle and D. Emmerick, J. Exp. Psych., 1966, 71, 794.

33. M. Levine, J. Exp. Psych., 1966, 71, 331.

34. J. Chumbley, J. Math. Psych., 1969, 6, 528.

35. L.E. Bourne Jr, Psych. Rev., 1970, 77, 546.

36. L.E. Bourne, Jr,"An inference model for conceptual rule learning", in Theories in cognitive psychology: the Loyola Symposium, Polomac, Md: Erlbaum, 1974.

37. R.L. Dominowski and N.E. Wetherick, J. Exp. Psych: Hum. Learn. and Mem., 1976, 2, 1.

38. J.S. Bruner et al., op. cit. (27), p. 22.

39. D.P. Ausubel, op. cit. (14), p. 510.

40. D.P. Ausubel, ibid., p. 517.

41. J.S. Bruner et al., op. cit. (27), p. 5.

42. A. Michotte, Act. Psych. 1950, 7, 298.

43. D.P. Ausubel, op. cit. (14), p. 511.

44. D.P. Ausubel, ibid., p. 53.

45. D.P. Ausubel, ibid., p. 509.

46. R.F. Kempa and G.H. Hodgson, Brit. J. Psych., 1976, 46, 253.

47. J. Cohen, Statistical power analysis for the behavioural sciences, Academic Press, 1969. Sections 3.3-3.4.

48. N.C. Kellett and A.H. Johnstone, Educ. Chem., 1974, 11, 111.

49. L.W. Gregg and H.A. Simon, J. Verb. Learn. and Verb. Behav., 1967, 6, 780.

50. B.R. Bugelski, J. Exp. Psych., 1962, 63, 409.

51. W. Kintsch, Learning, memory and conceptual processes, New York: John Wiley and Sons, 1970, Chapt. 4.

52. B. Milner, "Amnesia following operation in the temporal lobes" in Amnesia, O.L. Zangwill and C.M.W. Whitty (eds.), London: Butterworths, 1967.

53. M. Glanzer and A.R. Cunitz, J. Verb. Learn. and Verb. Behav., 1966, 5, 351.

54. E. Tulving and T.Y. Arbuckle, J. Verb. Learn. and Verb. Behav., 1963, 1, 321.

55. D.O. Hebb, "Distinctive features of learning in the higher animal", in Brain mechanisms and learning, J.F. Delafresnaye (Ed.), London: Oxford Uni. Press, 1961.

56. H. Buschke, J. Verb. Learn. and Verb. Behav., 1968, 7, 900.

57. G. Sperling, Psych. Monographs, 1960, 74, Whole Number 498.

58. G.A. Miller, Psych. Rev., 1956, 63, 81.

59. B.B. Murdoch Jr, J. Exp. Psych., 1961, 62, 618.

60. A.D. DeGroot, "Perception and memory versus thought: some old ideas and recent findings", in Problem solving, B. Kleinmuntz (Ed.), New York: John Wiley, 1966.

61. A.D. DeGroot, Thought and choice in chess, The Hague: Mouton, 1965.

62. G.V. Glass and J.C. Stanley, op. cit. (10), Sections 12.5 and 14.16.

63. D.A. Grant, Psych. Rev., 1962, 69, 54.

64. R.A. Fisher, Design of experiments, 8th Edition, New York: Hafner, 1960.

65. P.E. Meehl, Phil. Sci., 1967, 34, 103.

66. J. Cohen, op. cit. (47), p. 90.

67. J. Nunnally, Educ. and Psych. Measurement, 1960, 20, 641.

68. A. Binder, Psych. Rev., 1963, 70, 107.

69. J. Cohen, J. of Ab. and Social Psych., 1962, 65, 145.

70. J.K. Brewer, Am. Educ. Res. J., 1972, 9, 391.

71. W.W. Rozenboom, _Psych. Bull._, 1960, _57_, 416.

72. A. De Rujula, Invited talk on neutrino physics, Coral Gables Conference, Miami, 1976.

73. G.V. Glass and J.C. Stanley, _op. cit._ (10), Section 14.13.

74. G.V. Glass and J.C. Stanley, _ibid._, Section 11.3.

75. G.V. Glass and J.C. Stanley, _ibid._, Section 14.11.

76. J. Cohen, _op. cit._ (47), Sections 2.3.5 and 2.4.

77. _Structural Communications Topics_, University of London Press, 1970.

78. D.W. Massaro, _Experimental psychology and information processing_, Chicago: Rand McNally, 1975, pp. 250, 254 et. seq.

79. D.W. Massaro, _ibid._, p. 252.

80. A.D. Baddeley and G. Hitch, "Working memory", in _The psychology of learning and motivation: advances in research and theory Vol. VI_, New York: Academic Press, 1974.

81. E. Wanner and S. Shiner, _J. Verb. Learn. and Behav._, 1976, _15_, 159.

82. V.H. Yngve, _Proc. Am. Phil. Soc._, 1960, _104_, 444.

83. T.V. Howe, _Educational problems in writing chemical equations and formulae_, M.Sc. thesis, University of Glasgow, 1975.

84. I.M. Duncan and A.H. Johnstone, _Educ. Chem._, 1973, _10_, 213.

85. I.M. Duncan, _The use of programmes to investigate learning processes in difficult areas of school chemistry_, M.Sc. thesis, University of Glasgow, 1974.

86. F.M. Garforth, A.H. Johnstone and J.N. Lazonby, _Educ. Chem._, 1976, _13_, 41.

87. F.M. Garforth, A.H. Johnstone and J.N. Lazonby, _Educ. Chem._, 1976, _13_, 72.

88. F.M. Garforth, _Difficulties in the understanding and use of simple ionic equations in school chemistry_, B. Phil. thesis, University of Hull, 1975.

89. _Alternative Chemistry Syllabus - Memorandum No. 7_, Scottish Education Dept.

90. K. Urquhart, _Studies in the Chemistry/Mathematics boundary at secondary level_, M.Sc. thesis, University of Glasgow, 1975.

91. D.P. Ausubel, _op. cit._ (14), pp. 89-103.

92. D.J. Gunning and A.H. Johnstone, _Educ. Chem._, 1976, _13_, 12.