40414

National Library of Canada | Bibliothèque nationale du Canada

CANADIAN THESES ON MICROFICHE

THÈSES CANADIENNES SUR MICROFICHE

NAME OF AUTHOR/*NOM DE L'AUTEUR* ALAN KEITH GRIFFITHS

TITLE OF THESIS/*TITRE DE LA THÈSE* THE MOLE CONCEPT: INVESTIGATION OF AN HIERARCHICAL MODEL

UNIVERSITY/*UNIVERSITÉ* OF ALBERTA

DEGREE FOR WHICH THESIS WAS PRESENTED/ *GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE* Ph.D.

YEAR THIS DEGREE CONFERRED/*ANNÉE D'OBTENTION DE CE GRADE* 1979

NAME OF SUPERVISOR/*NOM DU DIRECTEUR DE THÈSE* DR H. KASS

Permission is hereby granted to the NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈQUE NATIONALE DU CANADA de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans l'autorisation écrite de l'auteur.

DATED/*DATÉ* 3rd December 1978 SIGNED/*SIGNÉ* Alan K. Griffiths

PERMANENT ADDRESS/*RÉSIDENCE FIXÉ* 03 O'REGAN RD
ST JOHN'S NEWFOUNDLAND
CANADA A1A 2J7.

NL-91 (3-74)

National Library of Canada

Cataloguing Branch
Canadian Theses Division

Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada

Direction du catalogage
Division des thèses canadiennes

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

LA THÈSE A ÉTÉ MICROFILMÉE TELLE QUE NOUS L'AVONS REÇUE

THE UNIVERSITY OF ALBERTA

THE MOLE CONCEPT: INVESTIGATION OF AN HIERARCHICAL MODEL

by

© ALAN KEITH GRIFFITHS

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF SECONDARY EDUCATION

EDMONTON, ALBERTA

SPRING, 1979

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled "The Mole Concept: Investigation of an Hierarchical Model" submitted by Alan Keith Griffiths in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

H. Kass
Supervisor

S. Hunka

L. E. McClung

R. G. Jachmicke

M. A. Nay

Glen S. Aikenhead
External Examiner

Date December 1st, 1978

ABSTRACT

The purposes of this study were threefold. First, it represents an attempt to validate an hierarchy leading to the learning of the concept of 'mole,' a concept which is generally of central importance in high school chemistry courses. Second, in order to achieve this first purpose it was necessary to examine the characteristics of the most promising methods available for the validation of hierarchies, both in terms of their theoretical bases and empirical relationships to one another. Third, the importance of the availability of subordinate skills within the validated hierarchy relative to the importance of learner developmental level to the acquisition of superordinate skills was examined.

The sample consisted of 269 Grade 10 Chemistry students in four Calgary senior high schools. Six test instruments were administered during the study, four of which represented the chemistry skills hypothesized to comprise the learning hierarchy leading to the mole concept. The remaining two tests were designed to elicit information about the prevailing stage of intellectual development of each individual in the sample. In addition two written units which were intended to remediate for subordinate skills which subjects failed to learn during regular instruction, were administered to a sub-sample of 109 subjects in four intact classes.

The hypothesized hierarchy was not found to be valid but an alternative hierarchy consisting of seven of the eight skills in the hypothesized hierarchy was considered valid, both in terms of the

psychometric and learning transfer relationships between the component skills.

Two recent psychometric methods due to White and Clark (1973) and Dayton and Macready (1976) appear to have substantial advantages over previous methods of hierarchy validation. Each of these was applied to the same data. A third method, the ordering-theoretic method, was also applied to the data for comparative purposes because of its recent prominence in the literature. It was concluded that the White and Clark test is more appropriate than the ordering-theoretic method in terms of its theoretical basis. However, the empirical results obtained from application of each method to the same data were very similar. It is suggested that in general the Dayton and Macready method is superior to the White and Clark test and the ordering-theoretic method because it considers the hierarchy as a whole rather than in terms of comparison of skills in pairs and also allows direct comparison of intact hierarchies. The hierarchy validated by the Dayton and Macready method was not identical to those derived from application of the other two methods.

Learner developmental level was found to exhibit only moderate correlations with achievement scores for the intellectual skills comprising the validated hierarchy. In all cases the availability of subordinate intellectual skills accounted for much more of the variance of scores on tests of related superordinate skills than did the developmental level test scores. The scores on the tests of related subordinate skills alone accounted for more than 50 percent of the variance of scores on the tests for each of the three uppermost skills in the validated hierarchy.

ACKNOWLEDGEMENTS

During the course of the study reported in this thesis I have benefited from the help of many people. In particular the editorial and conceptual guidance provided by my supervisor, Dr. Heidi Kass, has provided me with an invaluable model for the future.

Further thanks are gratefully accorded to the teachers, students and administrators of the Calgary Board of Education, without whose ready cooperation the study would not have been possible, and to my colleagues Dr. Alan Cornish and Mrs. Ruth Cornish for their help and advice with statistical and computational problems.

Last, but by no means least, the support of my wife Anne and my family, Anthony, Phillip and Kathryn, has sustained and inspired me throughout.

This research was partially supported by a Fellowship from Canada Council.

TABLE OF CONTENTS

APPENDICES PAGE

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

Chapter 1

THE PROBLEM

Introduction to the Problem

Educators, including those concerned with science education, have long wrestled with the problem of how to identify optimal sequences of instruction. In the first half of this century science education was much influenced by Dewey's (1916) suggestions for instruction sequenced to develop the ability to apply the 'scientific method.' More recently Schwab's (1966) emphasis upon the development of inquiry skills influenced curriculum sequences such as the 'BSCS Invitations to Inquiry.' Gagné's (1967) guidance may be seen in instructional sequences leading to the development of process skills in the 'Science A Process Approach' elementary science curriculum. The need to structure curricula according to their perceived major conceptual schemes (Bruner, 1960) is reflected in science curricula such as 'Chem Study,' the 'Biological Sciences Curriculum Study' and the 'Earth Science Curriculum Project,' as well as many others. Sequencing according to learner intellectual developmental levels is implied by Piaget (1964) and applied in curricula such as the 'Science Curriculum Improvement Study.'

Despite general agreement about the need to establish appropriate sequencing principles there is little agreement regarding what these should be. Briggs (1968) notes that the results of sequencing experiments are too scattered to allow confident recommendations for practice. Posner and Strike (1976) suggest that the literature indicates little more than that much more research is needed before we will be able to satisfactorily suggest how content should be sequenced. Indeed, they recommend that it is more appropriate first to obtain better answers to the descriptive question, 'In what ways can content be sequenced?' Posner and Strike's own theoretical and empirical analysis suggests five distinct categories of sequencing principle, namely world-related, concept-related, learning-related, inquiry-related, and utilization-related, each with a number of sub-categories. Researchers concerned with sequencing are likely to focus on one or other category or sub-category as the independent variable in their studies. The present study is involved with two aspects of sequencing from the learning-related category.

According to Berlinger and Gage (1976) the relationship between psychology and education is long but sometimes disappointing. Nevertheless, Berlinger and Gage go on to note that the marriage between the two seems stronger than ever. Shulman and Tamir (1973), reviewing science education of the previous decade, note the extensive influence of psychology on the development of science curricula. However, they caution that psychology is too frail a base to support an entire curriculum. In the same vein Shulman (1974) suggests that in education we should be concerned with middle-range theories, which lie between specific day to day working hypotheses and all-inclusive systematic efforts to develop a unified theory of curriculum and instruction. Shulman also suggests that the development of a psychology of school subjects should involve subject matter experts as well as psychologists rather than primarily psychologists as has typically been the case. The recent science education literature is consistent with the above views.

Different studies applying psychological theory to the sequencing problem as well as other aspects of curriculum and instruction reflect a broad spectrum of theoretical positions. The present study was conceived in an attempt to provide a partial answer to a persistent problem in high school chemistry, namely student difficulties in understanding the 'mole' concept. Gagné's contention that much of the curriculum content presented in schools may be best represented in learning hierarchies was considered a promising model from which to attack the mole problem. Essentially, Gagné suggests that the most important characteristic of the learner when he is faced with the learning and application of a new concept is that he should at the time be able to recall and apply all necessary directly prerequisite behaviors. For Gagné this is what constitutes 'developmental' readiness (Gagne, 1977, p. 150).

A different view of developmental readiness is taken by Piaget (1964). Piaget suggests that the most important determinant of an individual's readiness for new learning is whether he has developed appropriate generalized intellectual structures. Hence, even if a learning hierarchy can be identified for the mole concept, application of Piagetian principles suggests that individual progress through the hierarchy may be constrained by the intellectual limitations of the learner. Ausubel (1968) suggests that developmental readiness represents a composite of the positions taken by both Piaget and Gagné.

That is, he suggests that the availability of an adequately developed generalized intellectual structure as well as specific prerequisite competencies related to the particular subject matter are both important.

The present study seeks to attempt to identify an hierarchy for the mole concept, and also to determine if the developmental level of the learner, in the Piagetian sense, is related to the achievement of each part of the hierarchy. In the course of the study it became apparent that certain methodological problems exist for the determination of hierarchies. Research into some of these problems also became an important focus of the study.

What follows at this point is a summary of the relevant parts of the models provided by Gagné and Piaget, and a further discussion of the possibility of combining them. The chapter concludes with a discussion of the particular problem to be studied and the questions which it generates.

Gagné's Hierarchical Model of Learning

The content of this section represents but a small part of the contributions made by Robert Gagné to current views of learning. The titles of two of his major texts (Conditions of Learning (1965, 1970, 1977) and Essentials of Learning for Instruction (1974)) indicate the direction he has taken. From the first to the third edition of 'Conditions of Learning' many changes are apparent. For example the greater emphasis upon an 'information-processing' model in which interest centres on possible mechanisms by which input stimuli are transformed to output behaviors does not appear in the first edition, appears implicitly in the second, and becomes the explicit focus of the third edition. However, despite this and other significant changes, Gagné's emphasis upon hierarchies of learning remains an important feature throughout.

Despite his clear behaviorist origins, Gagné is careful to distinguish himself from the notion that any one prototype of learning is applicable to the domain of learning as a whole. He comments (Gagné, 1970):

> These learning prototypes all have a similar history in this respect: each of them started to be a representative of a particular variety of learning situation. Thorndike wanted to study animal association. Pavlov was studying reflexes. Ebbinghaus studied the memorization of verbal lists. Köhler was studying the solving of problems by animals. By some peculiar semantic process, these examples became prototypes of learning, and thus were considered to represent the domain of learning as a whole, or at least in large part. Somehow, they came to be placed in opposition to each other: either all learning was insight or all learning was conditioned response. Such controversies have continued for years, and have been relatively unproductive in advancing our understanding of learning as an event. (p. 20)

This belief in different types of learning led to the speculation that there might be as many different types as there would be different conditions under which learning takes place. Analysis suggested eight distinct types of learning (Gagné, 1965), and most importantly for the present discussion the suggestion was made that these types form an hierarchical arrangement in which successive types are prerequisite to the learning of the next. A slightly modified version of this model appears in Figure 1 (Gagné, 1970, p. 70). However, Gagné suggests that the four lower levels are likely to be applicable only

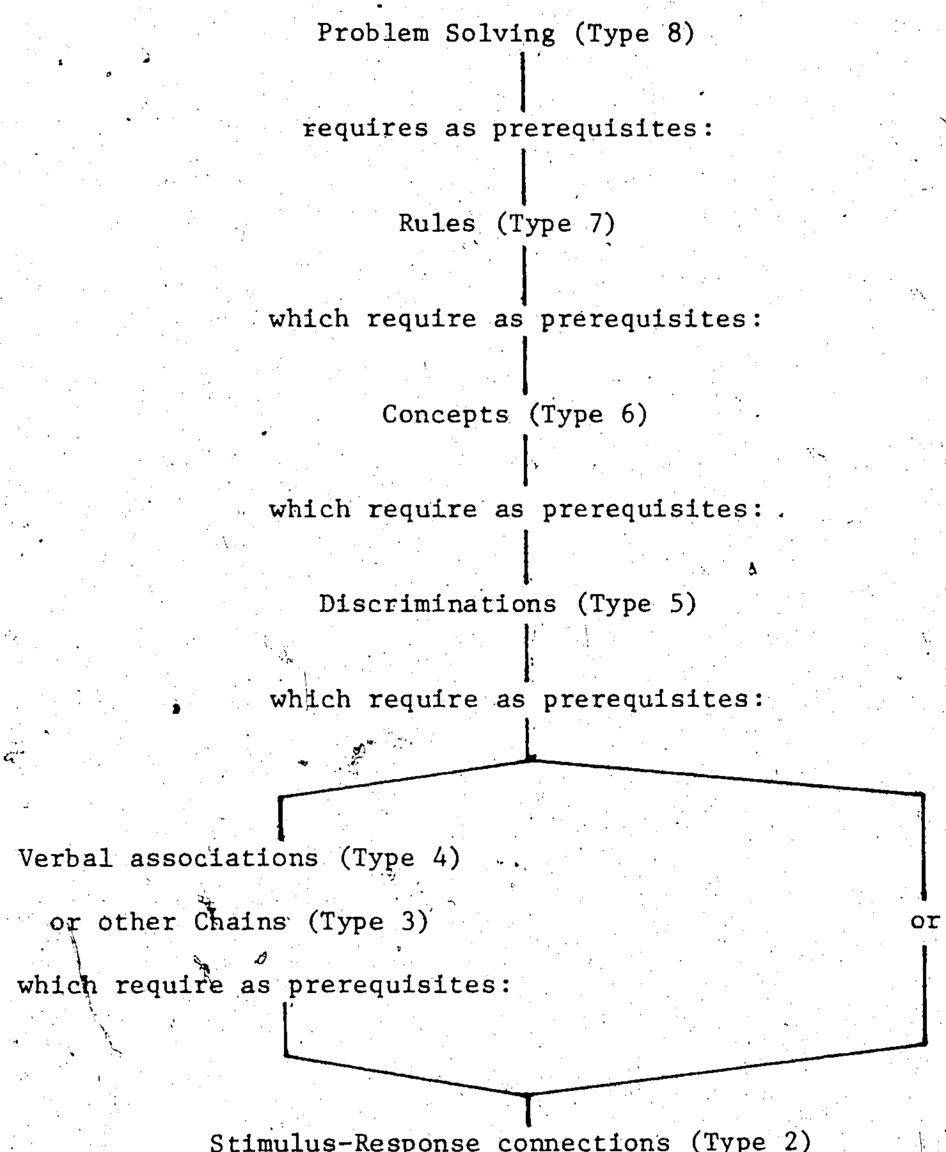


Fig. 1. Gagné's (1970) representation of learning types

to young children, and not of direct interest to those concerned with the instruction of older children.

Although Gagné's hierarchical model has been foundational to much of his work, a number of significant changes in his conception of the model may be seen since then. It is instructive to follow the development of the model and some criticisms of it. Of particular interest is the fact that while that part of the domain of learning to which Gagné suggests learning hierarchies to be applicable has diminished, the characteristics of hierarchies have become clearer. Correspondingly, this differentiation has clarified ideas about those learnings which do not appear to be representable by hierarchical structures.

An important question with which to begin relates to how learning hierarchies are generated. Gagné (1962) suggests that the components of the hierarchy should be generated by asking the question "What must the learner first be able to do if he is to achieve a particular new capability?" and then successively asking the same question for each new capability produced. The emphasis upon desired capabilities is important. Learning hierarchies represent what the learner should be able to do, not what he can merely verbalize. The resulting structure may be linear or branched such that in the latter case several capabilities may be considered directly prerequisite to any particular one. For example, it may be hypothesized that learners must first understand the concepts mass, volume, pressure and temperature if they are to exhibit understanding of the relationship conveyed by Boyle's law. However the crucial test of the validity of an hierarchy throughout Gagné's writings has been the extent to which learning of the prerequisites effects positive transfer to the learning of superordinate capabilities. That is, those who fail to exhibit prerequisite capabilities will fail to exhibit related superordinate ones while those who exhibit the prerequisites will be more likely to exhibit the superordinate capabilities. The corollary that increasing the proportion of learners who can exhibit the prerequisites will be followed by a significant increase in the proportion who can exhibit the superordinates, represents the essential test of the existence of an hierarchy (Gagné, 1970, p. 239).

This proposition is not entirely acceptable to others. White (1973) in particular has been severely critical of Gagné's index of positive transfer, and implicitly of the general notion of the importance of transfer. The relative merits of the 'psychometric' approach advocated by White, and of the transfer approach advocated by Gagné, will be discussed in the next chapter. Certainly, Gagné's emphasis upon transfer has not diminished although his methods of determining this effect have varied.

An important criticism of Gagné's procedure is made by Phillips and Kelly (1975) and reiterated by Posner and Strike (1976). Both suggest that Gagné conflates logical and psychological relationships when deriving his hierarchies. Posner and Strike distinguish between empirical prerequisites (a sub-type of their learning-related category of sequencing principle) and logical prerequisites (a sub-type of their concept-related category of sequencing principle) in that the latter is concerned with *a priori* properties of concepts while the former is concerned with skills that can be shown empirically to facilitate the learning of later skills. They further comment that what is logically necessary usually turns out to be empirically verifiable, but that the converse does not hold. While not denying the utility of a blend of these two kinds of relationships in curriculum materials, Posner and Strike suggest that it is important for the instructional designer to distinguish between them because logical prerequisites demand only logical analysis for justification while the existence of empirical prerequisites demands empirical evidence. Phillips and Kelly make the same point even more strongly, suggesting that Gagné's procedure for generating hypothesized hierarchies leads to hierarchies of primarily logical relationships within the subject matter. Hence, they argue, experimental validation by the determination of transfer relationships between prerequisite and superordinate behaviors is irrelevant because the 'discovered' relationships are inevitable, representing conceptual truths. Both Posner and Strike and Phillips and Kelly appear to be correct in suggesting that hypothetical learning hierarchies derived by the Gagnéan approach are likely to represent the logical structure of the subject rather than the psychological structure of the learner. However, conceptual truth for the neophyte learner is likely to be quite different than for the subject-matter expert. Perhaps this is why a reading of the literature indicates that expert-structured learning hierarchies do not survive empirical testing unscathed, and why Gagné's hierarchies exhibit a blending of logical and psychological relationships. Moreover, it may be argued that prerequisites which exhibit positive transfer may be considered to form part of the learner's psychological structure once learned, even if they were hypothesized by the researcher by logical analysis.

Reviews by Briggs (1968), Resnick and Wang (1969), Walbesser and Eisenberg (1972), White (1973) and others attest to the popularity of the learning hierarchy model. While the general conclusions of each of these reviewers are encouraging to the further application of the model, each notes the existence of many anomalies. Such anomalies have led Gagné (1972) to suggest that the hierarchical model may be restricted to certain kinds of learning. He postulates five domains of learning representing the learning of intellectual skills, cognitive strategies, motor skills, verbal information and attitudes, respectively. The emphasis here is still upon identifying different varieties of learning, and concomitantly the different conditions necessary for the learning of each variety. However, the eight types of learning posited originally (Gagné, 1965) now represent one of the five domains, the domain of intellectual skills. The hierarchical model remains (Gagné, 1977) as the most essential component of the conditions suggested for learning these skills. Gagné (1977) suggests that the empirical evidence available, while supporting the need to identify prerequisite behavior in any domain, does not at present support the suggestion that these prerequisites form a learning hierarchy in any domain other than the domain of intellectual skills. This claim is clearly of much importance to the present study. The nature of these domains and the restriction of the hierarchical model to the domain of intellectual skills is therefore explicated further, first by an illustrative example and then by further discussion:

Consider the variety of outcomes which may be involved when an individual is learning about Boyle's law. Correct application of the law in a meaningful way to a problem type which is not novel to the individual represents the use of an intellectual skill. The way in which the individual attacks a novel problem represents the use of particular cognitive strategies. A statement of the law itself represents the use of verbalized knowledge. Manipulation of apparatus in determining or applying the law involves the use of motor skills. Finally, the feelings of the individual as he does these things represent outcomes in the attitude domain.

The difference between intellectual skills and cognitive strategies may not be immediately apparent to the reader. In fact Gagné and Briggs (1974) describe cognitive strategies as very special kinds of intellectual skills, which are important enough and distinctive enough to warrant differentiation from other intellectual skills. Gagné and Briggs suggest that it is the object of the skill which differentiates cognitive strategies from other intellectual skills. Intellectual skills are oriented towards the individual's environment, such as the use of graphs or calculation of the number of atoms in a given mass of an element. In contrast, the essential characteristic of cognitive strategies is that they have as their objects the learner's own thought processes. Interestingly, given the joint focus of the present study, Gagné and Briggs suggest a link between the domain of cognitive strategies and Piaget's stages of intellectual development, in that they claim that the individual's repertoire of cognitive strategies sets limits to the kinds of problem solving he can successfully perform and hence to his intellectual development as defined by Piaget. Gagné and Briggs go on to imply that for Piaget the continuing development of cognitive structures is largely dependent on maturation, although Piaget (1964) has stated quite clearly that maturation is one of four factors determining intellectual development. According to Gagné and Briggs, development of cognitive strategies takes place by a process of generalization of more specifically learned intellectual skills. While it is not possible to indicate whether Piaget or Gagné is more correct in this respect, it should be acknowledged that Piaget's theory is much more rich in its description of the processes he believes to be involved. This contrasts with Gagné's (1977, p. 178) admission that "we do not yet know how to identify cognitive strategies, much less how to observe or measure them dependably." Appropriately, therefore, Gagné is not prepared to be very prescriptive when dealing with the development of cognitive strategies, and contents himself with recommending attention to the learning of basic intellectual skills and exposure to a variety of novel problems. This lack of prescription may be contrasted with Gagné's views about the development of intellectual skills.

Gagné and Briggs (1974, p. 49) suggest that while one can plan a sequence of events which will ensure the learning of intellectual skills, the design of instruction for cognitive strategies requires organization of events external to the learner so as to increase the probability of certain internal events. The design of instruction is in terms of 'favorable conditions' rather than 'sufficient conditions.' It is not to be expected therefore, if these suggestions are appropriate, that learning hierarchies leading to the learning of cognitive strategies will be identified.

The learning of a particular motor skill, according to Gagné (1977), also requires the recall of intellectual skills, and again does not appear to be representable by a learning hierarchy. The learner needs the availability of relevant intellectual skills in order to comprehend the executive sub-routine which defines the motor skill, and he is also likely to need to be able to exhibit the component parts of the skill, at least if the skill is not extremely simple. However, an individual who can exhibit each component part of a motor skill even at a low level of competency can usually also exhibit the complete skill to some extent. For example, consider a swimmer learning to crawl. Part skills would include particular kicking movements, arm actions, and head motions. There appears to be no particular sequence to the learning of these skills. On the contrary, all of them including the target skill are likely to be practiced simultaneously. Contrast this with learning the intellectual skill 'identifying horizontal and vertical components of forces as vectors' (Gagné, 1977, p. 144). Among other prerequisites this requires the prerequisite skill 'representing forces and their directions as parts of triangles,' which in turn requires the prerequisite 'identifying part of a triangle.' In this case the sequence of learning is clearly important. Exposure to the part skills is necessary before the individual can proceed to consider the overall skill.

The fourth domain, the domain of verbalizable knowledge, involves the learning of facts, statements of laws, etc. With regard to learning in this domain Gagné appears to have little to add to

Ausubel's (1968) writings. He agrees with the rote-meaningful learning distinction, and that for meaningful verbal learning to occur there is a need for an adequate existing base of related organized knowledge (Gagné, 1977). However, this base of knowledge may be quite idiosyncratic. Hence it appears that the contents of this domain also are not representable by learning hierarchies.

Finally, the fifth domain representing attitude formation also does not appear to be hierarchical in structure. According to Gagné (1974, p. 93) the main condition required for acceptable learning in this domain is exposure to a suitable model.

Each of the domains discussed above appears to demand different conditions for learning. Each is considered important in school learning. It is suggested, however, that substantial development in each of these domains demands the prior learning of relevant intellectual skills. The importance of the domain of intellectual skills is twofold. First, these skills are fundamental to learning in the other domains. Second, the domain of intellectual skills itself represents a substantial part of school learning. Hence the identification of hierarchies which are necessary to the development of these skills is of paramount importance. These hierarchies should contain only intellectual skills, and not elements from other domains. This does not suggest that the capabilities represented in the other four domains are irrelevant to the learning of intellectual skills, but only that they do not exhibit a consistent relationship to the intellectual skills in the hierarchy and cannot therefore be claimed to be a part of the hierarchy. In essence, learning hierarchies are not intended to offer a description of an entire instructional sequence. The suggestion made is that whatever else is present in the learning situation, the particular prerequisite skills must be available to the learner. The sequence represented by the learning hierarchy is not fundamentally immutable in a theoretical sense (Gagné, 1973), nor is it empirically immutable. Gagné (1970) describes it as an "on the average efficient route to the attainment of an organized set of intellectual skills." Resnick (1973) similarly refers to an hierarchy as the dominant sequence of acquisition of a set of intellectual skills.

Parallel to the development of a clearer definition of the domain to which learning hierarchies apply, a further significant limitation has appeared in Gagné's writings. This is with respect to the amount of content which should be considered as potentially hierarchical. Initially, it appears likely that Gagné considered not only that the whole domain of knowledge in any subject could be represented in an hierarchy, but that there was no limit to the size of hierarchy which might be constructed. In 1963 he commented (Gagné, 1963, p. 30),

> Knowledge is a hierarchy of ideas, in which the more complex ones depend for their acquisition of the previous mastery of simpler ones. Thus when a curriculum designer has in mind a set of ideas he wants students to acquire, he must ask himself very systematically, "What must the student already know how to do, in order that he can acquire this new knowledge?" The best way to construct a textbook, or an instructional sequence, is to begin at the end and work backwards with a rigorous application of this question at every step of the way. It is all too easy to construct inadequate sequences, and our present textbooks are full of them. Unless a student can progress regularly from one idea to the next we lose him.

Several years later (Gagné, 1967), the same view was still being expressed, perhaps even more strongly. A curriculum was defined as a "sequence of content units arranged in such a way that the learning of each unit may be accomplished as a single act, provided the capabilities described by prior units in the sequence have already been mastered by the learner." Clearly, such a curriculum has no upper or lower bounds. In the same publication a number of such curricula from quite diverse areas were cited. Generally, these were not complete curricula in the school sense, but they were certainly more extensive than the content of one lesson. Of interest to science educators, the kindergarten to grade six elementary science curriculum 'Science A Process Approach' was singled out as an especially significant example of the hierarchical model for curriculum development. There is no doubt that at that time Gagné saw this model as ~~of al~~most unbounded magnitude in curriculum development.

However, as Gagné has often indicated learning is affected by many factors, of which recall of prerequisite capabilities is but one part. Suggesting hierarchical connections on logical grounds is easier than validating them empirically. As will be indicated in the next chapter the appropriateness of the statistical and experimental methods used in many learning hierarchy studies may be challenged. Moreover, even if these methods were completely appropriate (and none are) it is clear that for larger hierarchies the many other factors which might affect learning are likely to complicate analysis of data designed to test the existence of those hierarchies. Whether for this reason or some other is not clear, but Gagné's more recent writings (Gagné, 1973) suggest that learning hierarchies are most clearly applicable to such components of curricula as single lessons. Certainly Briggs (1968), in reviewing the field, found it expeditious to focus upon brief units of instruction. Similarly, Glaser and Resnick (1972) note that carefully controlled studies of two or three-step transfer hierarchies have indicated significant positive transfer. However, the question of what constitutes an optimum amount of content for inclusion in a learning hierarchy is not yet answered. Expeditious as it may be for reasons of experimental control to focus upon very small hierarchies representing the content of one lesson, such hierarchies are unlikely to be considered of significant value to ducators. Conversely, it is here suggested that hierarchies representing several lessons, such as the hierarchy derived in Okey and Gagné's (1970) study, may allow sufficient experimental control and also represent an amount of content sufficiently large for the hierarchy to be useful in classroom practice. This is the position taken with respect to the content under consideration in the present study.

Piaget and Intellectual Development

Central to Piaget's theory is the view that the individual structures reality by a process of mental construction as he interacts with his environment. Intellectual development is the progressive differentiation of these mental structures from simple to complex. Basic to Piaget's theories is the belief that all individuals pass successively through several well defined stages of intellectual development, as inferred from the responses of subjects to a set of well defined tasks (Inhelder and Piaget, 1958). While the existence of intellectual stages has been strongly disputed (Brainerd, 1974) there is little question that an ultimate aim of education is to encourage development to behaviors of the kind represented by Piaget as formal operational. Typical manifestations of formal thought include systematic combination and control of a number of variables, interpretation of functional relationships, propositional reasoning, and proportional reasoning. Underlying these actions, according to Piaget, is the ability to apply the operations of the INRC group (inversion, negation, reciprocity, and correlation) (Inhelder and Piaget, 1958). The use of second-order operations, or combination of more than one operation, is another way of viewing formal operational thought.

A second key aspect of Piaget's theory is that although maturation, social interaction and experience all contribute to intellectual development, the single most important factor affecting this development is the occurrence of structural and functional equilibration. Karplus (1977) concisely describes equilibration as an internal mental process in which new experiences are combined with prior expectations to generate new logical operations. Essentially, the learner develops intellectual structures that enable him to assimilate new experiences, and he accommodates to these new experiences by construction of new intellectual structures.

These two aspects of Piaget's theory are particularly important to considerations of the content and experiences presented to the school learner. There is little doubt that a large proportion of high school students fail to show consistent evidence of formal operational thought in situations requiring it (Howe, 1974; Karplus et al., 1975). This is paralleled by suggestions that high school science content is often pitched at a formal operational level (Shayer, 1973; Herron, 1975; Karplus, 1977). Educators are therefore faced with a dilemma. Should the content of school curricula be pitched as nearly as possible to the intellectual level of the learner, or alternatively should it be pitched at a higher level in order to induce the learner to operate at this higher level. Educators are in dispute about the answer to this dilemma. Shayer (1973) suggests that the average high school student will not profit from exposure to formal operational content. He suggests biology and physics to be more appropriate than chemistry for the average high school student because these subjects as typically presented in high school are less intellectually demanding. Herron (1975) provides a list of chemistry concepts which may be accessible to the more concrete-operational thinkers, if presented only at a concrete level. Further, Herron (1976a) later suggests that many individuals in general chemistry courses need only a basic knowledge of chemistry without any depth of understanding, an example being students of home economics. Karplus (1977) argues that it is possible to lead the learner, through appropriate experiences, to self-regulation and towards extended formal thought. Thus, even given acceptance of Piaget's view that intellectual development precedes learning, these comments illustrate three different approaches to instructional design.

The Gagnéan and Piagetian Models: The Case for a Combined Approach

The preceding sections have indicated some important aspects of the writings of Gagné and Piaget as they relate to the development of concepts in school learning. It is possible to discern similarities and differences in the two models provided.

Strauss (1972) notes that Piaget and Gagné appear to agree that a child acquires an increasing number of intellectual capabilities as he grows older, that such acquisition is essential and that these capabilities are a product of the child's interaction with his environment. Strauss then goes on to argue that this apparent agreement is misleading, and that philosophically the two models are so disparate that attempts to combine them lead to obfuscations of the theoretical issues. With regard to the form of these intellectual capabilities, Strauss suggests that for Gagné they tend to represent narrowly defined structures, while for Piaget they are part of a much more comprehensive integrated structure. With regard to the particular prerequisites for the learning of a new concept Strauss appears to be correct in this interpretation. However, Gagné's claim that combination of a variety of intellectual skills enables the learner to generalize not only to form more complex intellectual skills but also to develop appropriate cognitive strategies, is a potential counter-argument to Strauss' claim.

A second difference between the two models, according to Strauss, is that in Gagné's model the learner is relatively passive and is copying rather than constructing reality, while according to Piaget the learner should be active in constructing his own intellectual structures. Again Gagné's distinction between the domain of intellectual skills and the other domains of learning, especially in this instance the domain of cognitive strategies, offers a potential answer to Strauss' objection. Gagné's emphasis upon the fundamental importance of intellectual skills leads to the need to carefully control the learning of those skills. In this sense Strauss is correct in suggesting that that the learner is copying reality. For Gagné, and for this writer, it does not make sense that the school learner, especially beyond the elementary grades, should form scientific concepts that are to any marked degree idiosyncratic. Neither does it seem likely that Piaget and his supporters would argue this point. The difference between Piaget and Gagné is more likely to centre around the degree of control imposed upon the learner.

One important aspect of this control is whether the majority of individuals need to follow the same instructional sequence to the learning of a particular concept as the learning hierarchy model suggests, or whether a variety of potential routes should be available. The difference between these two approaches represents a third point in Strauss' contention that the Gagnéan and Piagetian models are too far apart to allow meaningful combination. However, when the focus of interest shifts from the learning of a concept to its application to novel problems the differences become much less pronounced, as both Gagné and Piaget advocate that the learner should be active in constructing his own problem solving strategies.

Despite the very real distinction between the claims of Gagné and Piaget about the relationship between intellectual development and learning and also between the mechanisms they posit for how learning takes place, the models for curriculum and instruction derived from these theoretical positions may not be as disparate as Strauss suggests. The differences in theoretical positions need not preclude the application of either model to situations in which it may be of advantage. There are parallels to this in science. For example, the electron is alternately considered as a wave or a particle depending on the phenomena under consideration. Similarly, it may be important to consider the individual's understanding of prerequisite capabilities and his intellectual structures alternately as more or less important depending on the task at hand. There is support for this position in the educational literature.

Sticht (1971) suggests that within the cumulative Gagnéan learning model, ontogenetic development may set an upper limit to what can be learned at a given age. Correspondingly, he suggests that one troublesome problem within the Piagetian model may be related to a lack of subordinate competencies. This is the problem of horizontal décalage, a construct which Piaget uses to describe the inability of the individual to successfully perform a certain task despite the fact that he appears to have the necessary mental structures as evidenced by other related tasks.

The need to consider, for instructional purposes, the possibility of a profitable combination of the Piagetian and Gagnéan models may be further inferred from Beilin (1970). With particular reference to mathematics, Beilin warns of the danger of attempting to teach learning sequences derived *a priori*, and suggests the need for the use of both empirical and logical methods in the identification of hierarchical relationships between concepts and conceptual systems. His stricture that "logical relations are not inevitably paralleled by psychological relations" is of particular importance in the present context. Such lack of parallelism may be read in both directions, corresponding well with Sticht's comments. The need to investigate relationships between the conceptual systems of the subject and the cognitive system of the child is further stressed by Beilin. The position expressed by Sticht and Beilin is also often found in the writings of David Ausubel. For example:

> . . . it is legitimate to evaluate the internal logic of instructional materials from the standpoint of their appropriateness for learners at a specified level of intellectual ability and of subject matter and developmental readiness. (Ausubel, 1968, p. 29)

This position is foundational to the present study. In particular it is suggested that if it can be shown that hierarchical dependencies exist within typical high school science content, it does not seem appropriate to ignore them. Correspondingly, if it can be shown that developmental limitations might inhibit the learning of such content, it would not seem appropriate to ignore this evidence either. To put it more strongly it is suggested that it is inappropriate to ignore the evidence from either model in the application of the other.

Need for the Study

The science curriculum reform of the 1960's produced many changes in emphasis in the content of school science textbooks (Hurd, 1969).

One of these changes was towards a much greater emphasis upon the communication of the conceptual structure of the discipline. To this end major conceptual themes were identified by curriculum teams consisting of teachers, science educators and scientists. Text materials were then written in which these themes of important concepts formed a continuing structure throughout the text. With respect to chemistry curricula the materials produced by the writers of the internationally influential 'Chem Study' programme in the United States and the 'Nuffield O Level Chemistry' programmes in Great Britain were representative of this trend, and subsequent revisions have not changed the development of major conceptual themes as a basic emphasis. Examples of such themes include kinetic theory, equilibrium, the chemical bond, and energy.

As well as more faithfully representing the discipline, it was claimed that the development of concepts in integrated continuously evolving structures is an aid to learning and recall of these concepts (Bruner, 1960). However an inherent danger must be recognized, inasmuch as there is a possibility of cumulative failure when such important concepts are inadequately acquired by the learner.

It is suggested that the 'mole' represents an important conceptual theme in chemistry. Technically the mole is defined in its role as an SI standard unit of amount of substance as 'the amount of substance which contains as many elementary entities as there are carbon atoms in 0.012 kilogram of the carbon-12 isotope.' This number of elementary entities is, of course, equal to the Avogadro Number (L). In practice chemists extend the SI definition of mole so that a mole is the amount of a substance which contains L elementary entities, where the entity is specified to be an atom, a molecule, an electron, an ion, etc.

The usefulness of the mole to the student in an introductory chemistry course may be illustrated by reference to the following comment from the 'Handbook for Teachers' to the 'Nuffield O Level Chemistry' course (1967):

> The major, and almost the only reason for being concerned with weights of materials, is to put students as quickly as possible in a position in which what happens in a test tube becomes more intelligible because it can be discussed in terms of atoms, molecules, and ions. (p. 13).

The mole is the link which enables conversion between macroscopic quantities and the number or relative number of entities represented by these quantities. Such conversions are fundamental to an understanding and application of topics such as formulae of compounds, stoichiometric relationships, solution concentrations, electrochemistry, colligative properties, enthalpies of reactions, and a variety of equilibrium calculations.

In view of the foregoing comments it is not surprising that the mole is typically introduced early in high school chemistry texts. It is of concern that it is a source of difficulty to many students (Duncan and Johnstone, 1973; Novick and Menis, 1976). In a pilot study for the present research only seventeen percent of a sample of over two hundred grade ten chemistry students in an Edmonton high school were judged to have attained the mole concept after instruction, as evidenced by their ability to correctly answer questions such as 'What mass of calcium would have the same number of atoms as 9 grams of aluminum?' (molar masses were given).

An examination of high school chemistry materials indicates a variety of approaches in their treatment of the mole. Typically, the mole is referred to early, but the mode and depth of treatment varies considerably. In 'Nuffield O Level Chemistry' and also 'Chem Study' the treatment is early and relatively comprehensive. However, development of the mole in these programmes is in opposite directions. In Chem Study the order of treatment is from the gas laws to molar volume, then through Avogadro's Law to the Avogadro Number and hence to the mole. In the Nuffield programme the direction of treatment is reversed. A more recent programme 'Interdisciplinary Approaches to Chemistry' directed by Gardner (1973) deliberately offers an early but rudimentary introduction to integrative concepts such as the mole. The IAC authors assume that the learner will extend his understanding of these basic concepts as and when needed. Some authors recommend the application of algorithms such as the mole wheel (Head, 1968; Newstead, 1978; Ruda, 1978). Others suggest that such practices do not lead to understanding (Smith, 1978), a position with which the present author intuitively agrees. In short, the present position with respect to instruction of the mole and indeed to chemistry instruction in general, is that no coherent theoretical framework exists from which to make rational decisions. The models of learning and development provided by Gagné and Piaget offer potential solutions to this problem.

Within the Gagnéan framework learning of the mole concept represents acquisition of an intellectual skill. Therefore, it should be possible to identify a learning hierarchy which would provide an efficient route to the learning of the mole, such that learning of prerequisite skills in the hierarchy in the sequence identified will enhance the learning of the mole concept. Identification of such an hierarchy forms a major focus of the present study.

Analysis of the mole concept from a Piagetian perspective suggests that learner difficulty may be related to a mismatch between the intellectual competencies required to develop an understanding of the mole and the intellectual competencies available to the learner. Such a general lack of congruence between science curriculum materials and high school students has already been discussed. To illustrate this in the present context consider again the question 'What mass of calcium would contain the same number of atoms as 9 grams of aluminum?' Solution of this question requires the application of proportional reasoning and also the identification, recall and combination of several rules. For example, the learner may reason thus 'The mass of calcium containing the same number of atoms as 9 grams of aluminum would need to contain the same number of moles. Therefore I need to calculate this number of moles and then determine the mass of this number of moles of calcium.' Another method of solution might be to calculate the actual number of atoms involved, as an intermediate. In either case the application of proportional reasoning and second-order operations are involved, both aspects of formal operational thought in the Piagetian model. The relationship between the intellectual developmental level of the learner and acquisition of the mole concept is therefore a second focus of the present study.

As a result of the decision to apply Gagné's learning hierarchy model to the development of the mole concept, a further need which should be addressed in the present study was perceived. The review of the literature which follows in Chapter 2 will indicate extensive criticisms of the methods used to validate learning hierarchies. Several promising alternatives have recently been suggested. There is a need to examine these from a theoretical and empirical perspective, both as a contribution to the general research area and also because they appear to represent the most useful approaches to validation of the learning hierarchy which is proposed in the present study.

Purpose of the Study

The major purpose of the study is to identify a learning hierarchy for the mole concept as it appears in introductory high school chemistry. It is not suggested that it is possible to identify only one such hierarchy. What is suggested is that it is possible to hypothesize such an hierarchy by application of a task analysis procedure, and that this hierarchy or a modification of it may be validated empirically. This suggestion is consistent with Gagné (1973).

A secondary purpose of the study is to identify the extent of relationship between learner developmental level, as measured by selected tasks, and acquisition of the mole concept. The relative contributions of learner developmental level and availability of prerequisite skills to acquisition of the mole concept will be compared. Further, following earlier suggestions of possible combination of the Gagnéan and Piagetian models in the development of concepts the relationship between learner developmental level and acquisition of the prerequisite skills in any validated hierarchy will be considered.

A third purpose, at least as important as those already described, is to attempt to make recommendations relating to the theoretical and empirical significance of methods currently proposed for the validation of learning hierarchies.

Procedure

The procedure followed in the study is described in detail in Chapter 5. At this time it will be summarized briefly.

The sample, consisting of grade ten chemistry students, was assigned by intact classes to one of two groups, one of which was designated by the author to receive remedial instruction at appropriate pre-specified times. Several days prior to the commencement of treatment of the mole in their chemistry course all subjects responded individually both to a set of neo-Piagetian tasks and another test of intellectual development (the Skemp test), administered collectively to each class in a regularly scheduled double period. At appropriate times during the period of instruction of the mole each class wrote a total of three quizzes relating to those skills from the hypothesized hierarchy which had just been taught. At the end of the first two quizzes each subject in each class designated remedial was required to complete a written unit relating to the skills just tested as a homework assignment by the next scheduled class. Several days after completion of instruction of the mole concept all subjects wrote a final test composed of parallel items to those used in the quizzes.

Definition of Terms

Introductory chemistry: the beginning of formal chemistry in the senior high school. In the present study the text in use varied between schools, but was in all cases one of the versions of Chem Study.

Mole: the formal SI definition of 'mole' reads 'the amount of substance which contains as many elementary entities as there are carbon atoms in 0.012 kilogram of carbon-12' (Heslop and Wild, 1975). In common usage, and in the present study, this is extended to represent the mass of 6.03×10^{23} elementary entities such as atoms and molecules.

Possession of the mole concept: in the present study this is operationally defined as the ability to inter-convert masses of elements and compounds and the relative numbers of atoms or molecules contained in these masses, with at least seventy-five percent proficiency.

Operation: any representational act which is an integral part of an organized network of related acts (Flavell, 1963, p. 166).

Schema: an internalized mental representation of a particular kind of action.

For example, solution of one of the tasks used in this study requires the subject to apply proportional reasoning. If he has not developed the proportionality schema he is unlikely to be successful.

Developmental level: according to Piaget, four main stages occur in development. These are sensorimotor, pre-operational, concrete operational and formal operational. Piaget has described operations characteristic of each stage (Piaget, 1964).

Test of developmental level: a written test requiring some manipulations, composed of one item testing combination of variables (Hobbs, 1975), one item testing conservation of volume by displacement (Karplus and Lavatelli, 1969), and one item testing proportional reasoning (Wheeler, 1976).

Developmental level score: The total score obtained by an individual subject on the 'Test of developmental level.' Each item was scored 2, 1, or 0. The total score for an individual may range from zero to six.

Formal operational: an individual may be judged formal operational with respect to specific tasks, as well as with respect to a group of tasks. In general, the formal operational thinker can combine variables systematically, use proportional reasoning, conserve volume by displacement, realize the need to control variables systematically, and combine two or more operations. With respect to the present study an individual is judged formal if he scores five or six points on the 'Test of developmental level.'

Concrete operational: an individual may be judged concrete operational with respect to specific tasks, as well as with respect to a group of tasks. In general, he makes simple combinations of variables but does so unsystematically, uses an additive strategy where proportional reasoning is required, fails to conserve displacement volume, does not realize the need to control variables, and is not able to combine operations. In the present study an individual with a score of two or less on the 'Test of developmental level' is considered concrete operational. The assumption is made that none of the subjects in the study is below the concrete operational stage.

Intellectual skill: knowing *how* as contrasted with knowing *that* of information (Gagné, 1974a, p. 55). For example, how to calculate the mass of one mole of water, *not* that it is equal to 18 grams.

Learning hierarchy: use of the term 'learning hierarchy' is sufficiently varied to prevent statement of a generally accepted definition. However, the majority of researchers have used either a 'psychometric' or a 'transfer' definition, implicitly if not always explicitly. The essence of what is meant by a learning hierarchy may be understood more easily by reference first to a pair of skills which are hierarchically arranged. At the same time what is meant by the psychometric and transfer approaches may be illustrated. The psychometric definition is most clearly stated by White (1974). According to White the critical test of whether two skills are hierarchically arranged is that those individuals who can exhibit the superordinate skill are a sub-set of those who can exhibit the subordinate skill.

The transfer definition is well represented in Gagné's writings. According to Gagné (1970) two skills are hierarchically related if those individuals learning the subordinate skills have a significantly better chance of learning the superordinate skill. In this case, transfer of learning is said to take place from the subordinate skill to the superordinate skill.

In the present study each pair of skills involved in an hierarchy must be hierarchical when each of the above definitions is applied for the hierarchy to be considered valid.

A troublesome question in learning hierarchy research relates to the number of exceptions which can be allowed for an hierarchical connection between a pair of skills before the connection is declared invalid. Gagné's (1970) orientation to substantial rather than absolute hierarchical dependencies is applied in the present study. The imprecision implied by 'substantial' is acknowledged, but it is suggested that this lack of precision is unavoidable. This point is discussed further in Chapter 5 in connection with the methods examined in the present study.

Hypothesized hierarchy: a learning hierarchy composed of eight intellectual skills hypothesized by the author to be necessary for an individual to learn the mole concept. The skills are labelled A to H.

Final Chemistry Test: a written test compounded from eight criterion-referenced sub-tests of four items each. Each sub-test represents one of the intellectual skills in the hypothesized hierarchy. Skills G and H are tested only by multiple choice items. All other skills are tested only by free response items. The 32 items are scrambled throughout the Final test. No time limit was imposed, but the test was designed to be completed within one hour. Scoring criteria are discussed in Chapter 5.

Chemistry quizzes: three tests, between them testing the eight intellectual skills in the hypothesized hierarchy, containing parallel items to those used in the Final Chemistry Test. Quiz One tests skills G and H, while Quiz Two tests skills C, D, E and F, and Quiz Three tests skills A and B. No time limit was imposed in each case.

Remedial units: two written booklets, each containing further instruction and test questions representing selected intellectual skills from the hypothesized hierarchy. Remedial Unit One represents skills G and H. Remedial Unit Two represents skills C, D, E and F. Both units are reproduced in Appendices 8 and 9.

Delimitations of the Study

Restriction of the sample to one grade level (grade ten) in one Canadian city represents an important delimitation. Students of different chronological age, or who have different academic backgrounds might behave differently when instructed in the mole concept. The results obtained in the present study, while they may be considered suggestive, are not strictly generalizable beyond the present sample. It may be noted, however, that discussions with the participating teachers indicated that the sample represents a wide range of academic ability and socioeconomic status.

A further delimitation relates to the restriction to a selected area of the mole concept and a restricted set of subordinate concepts. The results obtained cannot be held to reflect upon the hierarchical nature of chemistry in general or of any other subject matter.

Finally, while the responses of each subject to the tasks involved in the Test of Developmental Level are assumed to be representative of the intellectual development of that individual, it is possible that other tasks would yield a different result. Nevertheless, within the limits of the time available for testing it is possible to apply only a selected representative sample of tasks from the domain of tasks available. This is typical of other related studies which have been reviewed by the investigator.

Limitations of the Study

Each of the quizzes and the Chemistry Final Test is composed of a set of sub-tests, each of which was constructed to represent a different intellectual skill from the hypothesized hierarchy. Each sub-test is considered to be a criterion-referenced test for the particular skill involved. The literature relating to criterion-referenced testing is reviewed later, but it may be noted here that there is dispute about what constitute appropriate measures of reliability and validity of such tests. It is not possible, therefore, to estimate the reliability and validity of the various chemistry sub-tests used in the present study by any uniformly agreed upon measures, and this constitutes a limitation to the use of these sub-tests.

A further limitation lies in the nature of learning hierarchy research in general. It is possible to identify an hierarchy which is useful in indicating that it appears to be necessary to learn some skills before others may be learned. It is also possible to identify an hierarchy such that some skills contribute to the learning of others. It is not possible to claim that the best hierarchy for either or both purposes has been identified. Nor is it possible to deny that some hierarchy may exist even if one is not found.

Research Questions and Hypotheses

The present study was conceived as an attempt to determine (1) if a learning hierarchy leading to acquisition of the mole concept may be identified and validated, and (2) if acquisition of the mole concept and the subordinate skills in a validated learning hierarchy culminating in acquisition of the mole concept are related to the intellectual development of those attempting to learn the concept. An hierarchy representing the mole concept was hypothesized and is presented in Figure 6 (p. 116).

It became apparent that a commonly accepted operational definition of what is meant by a 'valid' learning hierarchy does not exist. There are several reasons for this. First, some researchers focus upon a definition of a learning hierarchy for which acquisition of any skill which is superordinate to any other skill(s) in the hierarchy should be possible only for those learners who can exhibit the subordinate skill(s). This definition is often referred to as a 'psychometric' definition of what is meant by a learning hierarchy. Other researchers focus upon a 'transfer' definition of a learning hierarchy in which acquisition of the subordinate skill(s) should enable the learner to acquire related superordinate skills. In the ideal case there should be no exceptions to the hierarchy in the response patterns of a group of learners, and the same hierarchy would be 'validated' by both approaches. In practice, depending in part upon the purpose to which the hierarchy is to be applied, a proportion of exceptions is allowed which is not constant between different studies. Secondly, different psychometric tests and different tests of transfer have been applied in different studies. Thirdly, different degrees of statistical stringency have been applied in the use of these tests. What is meant when an hierarchy is said to be valid is therefore related to how the hierarchy has been validated and the stringency of the statistical tests applied. The term 'valid hierarchy' is therefore relatively arbitrary.

The problems of interpretation which are consequently raised are compounded by deficiencies in the statistical techniques which have been generally employed (White, 1973, 1974a). Recently several more promising alternative techniques have been suggested. These include the 'test of inclusion' (White and Clark, 1973), the 'ordering-theoretic' method (Bart and Krus, 1973) and a 'scaling' method (Dayton and Macready, 1976). It is desirable, therefore, to attempt to compare these techniques in terms of their theoretical bases and the empirical results which they provide. This forms an additional focus of the present study. As will be seen from the discussion of these methods in Chapter 2, the White and Clark test and the Dayton and

Macready method are considered the most suitable psychometric methods available at the present time. These methods will therefore be used to judge the validity of any hierarchy emerging from the present study. As will be seen from the discussion in Chapter 2 the ordering-theoretic method is less pleasing conceptually than the White and Clark test or the Dayton and Macready method, but is simpler to apply. For comparative purposes the results of applying the ordering-theoretic method and the White and Clark test to the same data will be considered. Finally, any 'validated' hierarchy representing the mole concept which emerges after application of the White and Clark test and the Dayton and Macready method to the data will be further evaluated by a 'test of transfer' developed within the context of the present study.

The foregoing discussion leads to a number of research questions which will now be stated:

Question 1: Does the arrangement of intellectual skills represented in the hypothesized hierarchy represent a learning hierarchy which is valid according to the application of the White and Clark test?

If the answer to question one is negative question two will be considered.

Question 2: Does some other arrangement of some or all of the intellectual skills represented in the hypothesized hierarchy represent a learning hierarchy which is valid according to the application of the White and Clark test?

Because they both test the relationship between pairs of skills, and given that both are currently popular in the literature the relationship between the hierarchical connections identified by the ordering-theoretic method and the White and Clark test is of interest. This relationship is addressed in question 3.

Question 3: What is the extent of agreement between the results of applying the White and Clark test and the ordering-theoretic method to the same data?

The fourth research question concerns the validity of the hypothesized hierarchy and modifications of it according to Dayton and Macready's criteria.

Question 4: Does the arrangement of intellectual skills represented in the hypothesized hierarchy or an alternate arrangement of some or all of these skills represent a learning hierarchy which is valid according to the application of the Dayton and Macready method?

The fifth research question relates to identification of transfer of learning within any hierarchy considered valid after answering questions one to four.

Question 5: Is significant positive transfer (as evidenced by the 'test of transfer') observed between subordinate skills and related superordinate skills in the validated hierarchy?

The final two questions relate to the relationship between the intellectual developmental level of the learner and his acquisition of the skills in the validated learning hierarchy. Two independent tests of the learner's intellectual level of development are applied, the 'Test of Developmental Level' which is represented by a group of three tasks selected for the present purpose and the 'Skemp' test. Continuous scores for each are applied separately in addressing questions 6 and 7. Question 6 relates directly to the relationship between developmental level and acquisition of the skills in the validated hierarchy. Question 7 relates to the relative relationship of developmental level and availability of subordinate skills as predictors of achievement of the superordinate skill in the validated hierarchy.

Question 6: Does a significant relationship exist between developmental level and acquisition of each skill in the validated hierarchy?

Question 7: What is the relationship between developmental level scores and scores on the tests for subordinate skills as predictors of achievement of the superordinate skills in the validated hierarchy?

The statistical tests involved in answering questions 1, 2 and 4 are implied within the methods to which these questions refer. The answers to the remaining four questions may be tested statistically. Hence each is restated as an hypothesis. Hypotheses 1, 2, 3 and 4 are derived from research questions 3, 5, 6 and 7, respectively.

Hypothesis 1: No significant agreement exists between the skills identified by application of the White and Clark test and the ordering-the c method to the same data.

Hypothesis 2: No sign icant relationship exists between acquisition of subordinate sk lls and acquisition of related superordinate skills in the validated hierarchy.

Hypothesis 3 (a): No significant relationship exists between achievement on the 'Test of Developmental Level' and achievement on the tests for each skill in the validated hierarchy.

Hypothesis 3 (b): No significant relationship exists between achievement on the Skemp test and achievement on the tests for each skill in the validated hierarchy.

Hypothesis 4: No significant difference exists between achievement on the 'Test of Developmental Level' and the Skemp test relative to achievement of subordinate skills in the validated hierarchy as predictors of achievement of the superordinate skill in the validated hierarchy.

Summary

The possibility of applying Gagné's hierarchical model of learning to the learning of the 'mole' concept, typically encountered as an important integrating concept in introductory chemistry courses, has been discussed. Limitations in the intellectual development of the learner have also been suggested as exerting a potentially limiting influence on the learner's ability to acquire the mole concept. The possible interaction of Gagné's learning hierarchy model and Piaget's developmental model of intellectual functioning to instructional problems such as the learning development of the mole concept has been discussed, following a brief summary of Gagnéan and Piagetian theories as they apply to instruction. Procedures for identifying a learning hierarchy leading to the acquisition of the mole concept and for identifying the relationship between acquisition of the skills represented in the learning hierarchy and learner developmental level have been indicated briefly, followed by the presentation of definitions, delimitations, limitations and the research questions and hypotheses which are involved in the study. The chapter concludes with an overview of the report.

Overview

One of the problems encountered in the determination of learning hierarchies relates to the use of different methods and criteria for validation. The next chapter involves a description and discussion of the essential features of the most important techniques which have been used to identify learning hierarchies, and considers in more detail several recent methods which were applied in the present study. Chapters three and four describe empirical studies relating to hierarchies in science instruction and to the relationship between developmental level and achievement in science, respectively. Chapter four concludes with an examination of the evidence for group-testing techniques for the identification of developmental level, a technique that was applied in the present study. The design of the study and a description of the test instruments and procedures used are presented in Chapter five. Chapter six describes the analysis of data and the results obtained from the study. The report concludes with Chapter seven which includes a summary of the study and the major conclusions and recommendations for further research.

Chapter 2

METHODS USED TO VALIDATE LEARNING HIERARCHIES

Introduction

The methods which have been used to validate learning hierarchies fall into three main categories. These may be represented as involving transfer, psychometric techniques involving scaling, and psychometric techniques based upon the use of contingency tables.

Paradoxically, in addressing the methods which have been used to determine the validity of learning hierarchies it is not possible to specify precisely what constitutes a valid hierarchy. The general form of a learning hierarchy as a network of connections between pairs of intellectual skills is clear. Ideally, if a psychometric test of the validity of an hierarchy is applied, there should be no exception to the superordinate-subordinate character of each connection in the hierarchy. Similarly, if the test applied is of the transfer type all individuals who learn the subordinate skills in the hierarchy should learn the superordinate skill. However two problems complicate these ideal situations. Firstly, most investigators are prepared to modify the ideal relationships described above to allow some exceptions. Hence, substantial rather than perfect hierarchical dependence is implied. Operationally, the meaning of substantial varies according to the perspectives of different investigators. Secondly, except in the ideal case described above, an hierarchy declared valid either according to a psychometric test or a transfer test will not necessarily be valid according to the other. Moreover, the application of different psychometric tests or different transfer tests may also lead to different judgments of the validity of an hierarchy under test.

In practice these problems are not always severe. When there are few exceptions to an hierarchy, different methods yield similar results. Moreover, recent conceptual and empirical advances with respect to the validation of learning hierarchies offer much promise. Hopefully, the present study will contribute to this development. The review which follows in the remainder of this chapter is intended to indicate the most important methods which have been used to validate learning hierarchies, and to provide a basis for the methods used in the present study. In this chapter emphasis will be placed upon the methodology of hierarchy validation. The substantive aspects of some of the hierarchies derived from studies discussed in this chapter, as well as those derived from other selected studies, will be discussed in the next chapter.

Gagné's Index of Positive Transfer and Other Related Indices

The foundational and continuing influense exerted by Gagné in the development of the learning hierarchy model merits consideration of his methods of validation.

The foundation of the learning hierarchy model was laid in a series of studies conducted by Gagné and his colleagues in the early nineteen sixties (Gagné, 1962; Gagné and Paradise, 1961; Gagné, Mayor, Garstens and Paradise, 1962). Each of these studies involved the learning of mathematics skills, and in each case the skills formed a sequence in which it was not generally found that an individual learned a later skill in the sequence unless he had previously learned all of the earlier skills. It was suggested therefore that the learning of earlier skills enhanced the learning of each skill occurring later. From this Gagné derived the concept of positive transfer and an index of proportion positive transfer was derived. This index may be illustrated by reference to Figure 2, in which the letters A to D reflect the frequencies of subjects exhibiting the particular pass-fail relationships described below:

A = number of subjects who achieve the upper skill and all related lower skills

B = number of subjects who achieve the upper skill but fail to achieve at least one related lower skill

C = number of subjects who fail to achieve the upper skill and at least one related lower skill

D = number of subjects who achieve the upper skill and fail to achieve all related lower skills

In each case the number of related lower skills refers to those immediately subordinate to the upper skill in question. According to Gagné responses A and C are supportive of the hypothesized transfer effect, while responses D are contrary to it. Responses B are considered neutral. The index of proportion positive transfer was therefore defined as the number of supporting responses divided by the number of supporting and contrary responses taken together. This is represented in the expression:

$$\text{Proportion position transfer} = \frac{A + C}{A + C + D}$$

The index was set at a criterion level of 0.90, and the validity of

Lower skill(s)		Upper skill: Fail	Upper skill: Pass
	Pass	B	A
	Fail	C	D

Fig. 2. Contingency table for calculation of proportion positive transfer

the hierarchy judged accordingly.

The index of proportion positive transfer is conceptually pleasing but practically limited. White (1973) demonstrates clearly that many of the apparently acceptable values cannot be distinguished from chance. He also (appropriately) comments that a high value for this index may merely reflect a positive correlation, representing a necessary but not sufficient criterion for an hierarchical relationship. However, the usefulness of the *concept* of positive transfer is not this easily dismissed. It may certainly be addressed in other ways, for example by comparing the achievement of a group taught through the hierarchy with a group not so taught (Okey and Gagné, 1970). It may be argued that the efficiency of a learning hierarchy in promoting transfer represents the crucial question for research once the existence of an hierarchical sequence has been established. Gagné (1974) continues to emphasize its importance. Carroll (1973) considers it to be the essential criterion of the validity of a learning hierarchy. Phillips (1974) notes that in general the notion of positive transfer through provision of learning hierarchies is supported by substantial evidence. Nevertheless, White's contention that the index of proportion positive transfer is not a useful measure of the validity of an hierarchy is accepted by the present investigator.

Eisenberg and Walbesser (1971) suggested the use of a further series of indices based upon the frequencies represented in Figure 2 as measures of hierarchical dependency. Five indices were developed:

1. *A Consistency Ratio*, testing the implication that acquisition

of the terminal behavior implies acquisition of all subordinate behaviors.

$$\text{Consistency} = \frac{A}{A + D}$$

Clearly, the ratio is meaningful in that it reflects the strength of the hierarchy. A low value destroys the notion of a viable sequence.

2. *An Adequacy Ratio*, testing the implication that acquisition of all subordinate behaviors implies, with adequate instruction, acquisition of the terminal behavior:

$$\text{Adequacy} = \frac{A}{A + B}$$

3. *An Inverse Consistency Ratio*, meant to test the implication that non-acquisition of the terminal behavior implies non-acquisition of all subordinate behaviors, represented by

$$\frac{C}{C + B}$$

4. *An Inverse Adequacy Ratio*, testing the implication that non-acquisition of all subordinate behaviors implies, even with mediating instruction, non-acquisition of the terminal behavior, represented by

$$\frac{C}{C + D}$$

These last two ratios appear to be quite inadequate in themselves because an unsuccessful hierarchy or unsuccessful instruction or a combination of these would yield a high (C,C) loading and a low loading in the other cells. In each instance a very high value of the index would be obtained. In addition, the author's claim that the two ratios cannot be high simultaneously would be incorrect in this instance.

5. *A completeness Ratio*, which focuses on the proportion of individuals capable of successfully completing the whole sequence, is defined as

$$\frac{A}{A + C}$$

Walbesser and Eisenberg suggested arbitrarily that for a learning hierarchy to be considered valid values of at least 0.85 should be obtained for the consistency, adequacy and completeness ratios for each connection in the hierarchy. Moreover, the other two ratios they described were suggested to provide useful information to the investigator. Once again White (1974) was able to produce perceptive and appropriate criticisms. White showed that for the three indices recommended by Eisenberg and Walbesser to have a minimum value of 0.85, more than 65 percent of the subjects must achieve the highest skill in the hierarchy. Hence nearly all subjects will be achieving the lower skills and it is impossible to say that subjects failing to learn the lower skill cannot learn higher ones because such subjects are not present. Application of Eisenberg and Walbesser's indices to the data from three of Gagné's studies supported only five out of 31 connections found by Gagné.

Capie and Jones (1971) applied Eisenberg and Walbesser's consistency and adequacy ratios to the determination of an hierarchy leading to the application of specific gravity rules. Their main purpose, however, was to compare the application of different hierarchy validation methods to the same data. As well as the two indices referred to they added a 'necessity' ratio, defined as

$$N = \frac{C}{C + D}$$

in terms of the frequencies in Figure 2, and also the calculation of a phi correlation coefficient between subjects' responses to adjacent skills. The results of applying these different measures to the same data were very inconsistent.

Two further comments regarding the measures used by Capie and Jones may be made. First, the advantages of the necessity ratio are not clear. It was claimed to be particularly useful when the subjects represent a wide range of ability. However, the highest possible value for N would be when all subjects failed both skills, at best minimally positive evidence for the hierarchy. Secondly, Capie and Jones noted that the pattern of decreasing phi correlations between the skills in their hypothetical hierarchy as these skills became further removed was consistent with the existence of a valid hierarchy. White (1974), however, notes that while such a result is consistent with an hierarchical relationship it is not sufficient in itself to support the validity of the hierarchy.

Each of the indices described in this section is considered to be of very limited usefulness in the validation of learning hierarchies. White in particular has been critical of their value. He summarizes their limitations (White, 1974) as follows:

> All are shown to be unsatisfactory for one or more of the following reasons: sometimes the index can have values which indicate a hierarchical connection even when the skills are really independent; errors of measurement are ignored; there is no way of calculating the probability of error in generalizing from sample to population; and subjective judgements are necessary on how many questions

of a set must be answered correctly for a "pass" in a skill.

In the same paper for a given set of real data White showed that Walbesser's indices nearly always lead to acceptance of connections and proportion position transfer to a state of indecision. Such criticisms are potentially devastating to the credence researchers can give existing findings. However, in order to be constructive, some better measure which takes account of at least some of these criticisms must be devised. White and Clark (1973) had already attempted to do this. The method which they developed represents the focus of the next section.

The White and Clark Test of Inclusion

The White and Clark test focuses upon comparison of pairs of skills in the hierarchy. It allows for errors of measurement and also provides a test of statistical significance. Subjects are required to answer at least two questions per skill. In fact, White and Clark offer solutions only for the cases when both skills are tested by two questions or by three questions. (Griffiths and Cornish (1978) have extended this to generalize to any number of questions per skill, and for unequal numbers of questions per skill. They also make appropriate modifications to the significance test.) A skill-by-skill matrix of scores is then formed, the scores ranging from zero to two or more as appropriate. The matrix representing two questions per skill is shown in Figure 3.

The cell representing a score of zero on the lower skill and maximum possible on the upper skill is used to test the hierarchical

Skill II (Upper) Questions Correct

Skill I (Lower) Questions Correct

	0	1	2
2			
1			
0			X

The critical cell is marked X

Fig. 3 Data matrix for the White and Clark Test relationship. This cell is assumed to contain those subjects most likely possessing the upper skill and lacking the lower one. The basis of the method is to test the null hypothesis that there will be no entries in this critical cell other than those representing errors of measurement. The probability that the observed frequency does not violate the null hypothesis is calculated by using the marginal totals. For the case of two questions per skill the probability that a member of the sample will be found in the critical cell is:

$$P_{02} = P_0\ (1-\theta_b)^2\ \theta_d^2 + P_I\ (1-\theta_a)^2\ \theta_d^2 + P_{II}\ (1-\theta_b)^2\ \theta_c^2 + P_B\ (1-\theta_a)^2\ \theta_c^2,$$

where

P_0 = the proportion of the population with neither skill

P_B = the proportion of the population with both skills

P_I = the proportion of the population with skill I only

P_{II} = the proportion of the population with skill II only

θ_a = the probability of someone with skill I answering correctly any skill I question

θ_b = the probability of someone without skill I answering correctly any skill I question

θ_c, θ_d are the corresponding probabilities for skill II

To make the estimate for P_{02} as large as possible and hence reduce the possibility of Type I error, θ_b is assumed to be zero and θ_c is assumed to equal one. That is, it is assumed that all subjects with one skill I question correct really possessed skill I and all those with one skill II question correct lacked the skill. For the case of three questions per skill appropriate modifications are made to the derivation above. In each case the hierarchical nature of all pairs of connected skills in an hypothesized hierarchy is tested, and the validity of the composite hierarchy then judged.

The White and Clark test represents a psychometric approach to hierarchy validation. It should be reiterated that an hierarchy derived in this way does not necessarily correspond to a transfer hierarchy leading to the learning of the same superordinate skill. An hierarchy validated by application of the White and Clark test may be said to represent an absolute level of hierarchical dependence. As has been noted Gagné suggests that substantial rather than absolute hierarchical dependence may be more reasonable. This suggestion has been adopted by Linke (1975) who has modified the White and Clark test to allow the same procedures to be used with one and two percent exceptions in addition to those representing errors of measurement, and by Beeson (1977) who similarly allows for five percent exceptions.

Another method which focuses upon pairs of skills has been described by Bart and others. It is described in the next section.

The Ordering-Theoretic Method

Bart and Krus (1973) and Airasian and Bart (1975) propose the application of what they call an 'ordering-theoretic method' to the validation of hierarchies. The method may be described by reference to Figure 2 (p. 46) with slight redefinition of terms to allow for the fact that skills are compared in pairs and hence there is only one lower skill in each test of an hierarchical connection. The method focuses upon the percentage of subjects whose responses are disconfirmatory to the existence of an hierarchical connection between two skills. In terms of Figure 2 the critical test is that the frequency represented by D should not exceed a prespecified tolerance level. The test is applied to all possible combinations of pairs of skills in the hierarchy, including those not hypothesized to be hierarchical. A composite hierarchy is then identified. The method is deterministic, and does not take into account errors of measurement. No test is provided to determine the statistical confidence which can be attached to the existence of each accepted hierarchical relationship, although a test by Wood (1975) is suggested to be useful in determining whether the number of hierarchical connections obtained exceeds what could be expected by chance. However Cotton, Gallagher and Marshall (1977) note that Wood's test is biased against confirmation of hierarchical connections when the number of possible connections is small.

Airasian and Bart (1975) illustrate the usefulness of the ordering-theoretic approach in identifying unexpected hierarchical relationships by applying it to data from Gagné and colleagues' (1962) seminal study. The ordering-theoretic approach appeared to identify a number of hierarchical relationships additional to those found in the original study. However, the authors appear to have missed the major thrust of the original paper, namely that the hierarchy should be defined in terms of transfer of learning.

A further problem likely to confound the results of applying ordering-theory is described by Wellens, Lenke, and Oswald (1977). These authors note the current unresolved debate about the assessment of appropriate cut-off scores for mastery and show that adopting different 'recommended' criteria for mastery may result in quite different hierarchies. While this is also true of many other methods it does not apply, for example, to the White and Clark test.

Bart and Krus claim that the ordering-theoretic method is superior to 'scaling' methods derived from Guttman (1944) in that such methods are restricted to linear hierarchies or composite hierarchies derived from linear combinations of skills. As will be seen in a later section this claim is not true of a very recent scaling method derived by Dayton and Macready (1976). An extensive discussion of the Dayton and Macready method follows the next section. The next section itself is concerned with a summary of Guttman scaling as it applies to hierarchy validation.

Guttman Scaling and Some Modifications

To illustrate what is meant by a Guttman scale consider a test or questionnaire in which there is one acceptable response to each item. When the responses of a number of individuals to these items form a sequence such that all individuals who make an acceptable response to any particular item also make an acceptable response to all earlier items, the sequence represents the ideal form of what Guttman (1944) called a scale. An hierarchy which is linear is clearly an example of a Guttman scale.

In practice the responses of some subjects will not conform perfectly to the scale. Such responses constitute error. To maintain some standard for acceptance or rejection of the scale, and at the same time allow for some reasonable level of error, Guttman derived an index of 'reproducibility.' This was defined as the quotient of total errors (i.e. deviations from a perfect scale) over total responses, subtracted from one. Arbitrarily, a reproducibility (REP) of at least 0.90 was declared necessary if the hypothesized scale was to be considered valid (Guttman, 1944).

When applied to the validation of learning hierarchies Guttman scaling has an obvious limitation, in that it can only be applied to linear hierarchies. An extension of Guttman scaling to allow for the identification of sub-scales which could then be linked together to form a branched hierarchy was developed by Lingoes (1963), and was termed multiple scalogram analysis. Both Guttman and Lingoes' scaling methods have been applied to hierarchy validation, and in the next chapter several such studies which are relevant to the present study will be discussed. Lingoes' method will not be described in more detail because it is not central to the development of the present discussion.

As well as being limited to linear scales or combinations of linear scales, the Guttman and Lingoes' scaling methods are further restricted in usefulness in that the 'statistical' tests which they use to judge the validity of a scale or hierarchy were suggested by their authors on arbitrary grounds. Proctor (1970) suggests them to be pre-statistical, and describes a method designed to elevate Guttman scaling to a better statistical foundation.

The basis of Proctor's model may be illustrated by the following example. Suppose for a three-item scale every subject in a population belongs to one of four Guttman true types which we denote by the vectors (0,0,0), (1,0,0), (1,1,0), (1,1,1), (where 0,1 represent non-possession and possession, respectively, of a skill and the hypothesized lowest skill is the first element and the highest skill is the last element in each vector). In a test of the validity of the scale there will likely be subjects whose response patterns are not one of the four valid ones above. Under the null hypothesis such invalid responses are misclassifications. Proctor uses a misclassification parameter α to represent the probability of an invalid response to an item. If $\theta_1, \theta_2, \theta_3$ and θ_4 are the proportions of the true types in the population then the probability of a response of the type (0,1,0) for example is given by

$$Pr(0,1,0) = \alpha(1-\alpha)^2\theta_1 + \alpha^2(1-\alpha)\theta_2 + \alpha(1-\alpha)^2\theta_3 + \alpha^2(1-\alpha)\theta_4$$

Assuming a multinominal distribution of frequencies of the observed response patterns for a given sample, Proctor then obtains maximum likelihood estimates of the proportions of true types and the misclassification parameter by an iterative procedure, and determines the goodness of fit between data and model by a chi-square test. Proctor's suggestion that his model is of intrinsic interest is modest Although it has not been directly applied in validation of learning hierarchies, it forms the basis for Dayton and Macready's (1976) attempt to overcome the other main objection to Guttman scaling. That is, the Dayton and Macready method offers the possibility of extension to hierarchies of any configuration. These authors also capitalize on a suggestion by Proctor that it should be possible to allow separately for a '1-for-a-0' error and a '0-for-a-1' error. These they call 'guessing' and 'forgetting' parameters, α and β respectively.

The Dayton and Macready Model

Dayton and Macready's extension is discussed in some detail because it is one of the methods used in the present study. Consider, for example, the case of a five element hierarchy. Whatever the configuration of the hierarchy there are 32 possible distinct response patterns. If the hierarchy is linear only six of these are in agreement with the hierarchy. The remaining 26 would be considered misclassifications.

Now consider the two branched hierarchies represented in Figure 4, each containing five elements. In (a) ten response patterns (00000), (10000), (11000), (00100), (10100), (11100), (00110), (10110), (11110), and (11111) are true to the hierarchy, with the remaining 22 patterns representing errors. In (b) there would be only seven true response patterns, namely (00000), (10000), (01000), (11000), (11100), (11110), (11111).

Using Dayton and Macready's (1976) notation, the probability P(u) of a subject producing a specific response pattern 'u' if the hierarchy is valid, is given by

(i) $$P(u) = \sum_{i=1}^{q} P(u|\nu_j)\,\theta_j$$

where ν_j represents the set of q true response patterns, and

(ii) $$P(u|\nu_j) = \prod_{i=1}^{k} \alpha_i^{a_{ij}} (1-\alpha_i)^{b_{ij}} \beta_i^{c_{ij}} (1-\beta_i)^{d_{ij}}$$

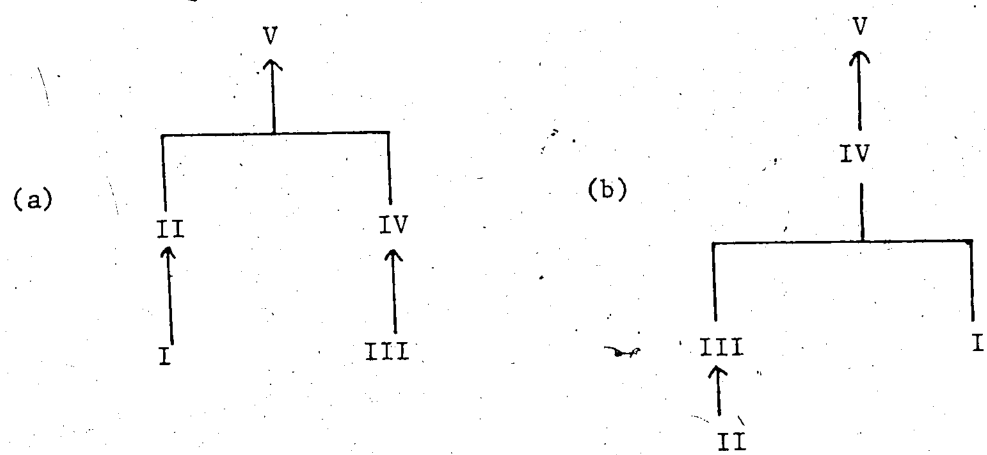


Fig. 4. Sample hierarchies for the Dayton and Macready Model

Essentially what is implied by the use of these two equations is that the values of the misclassification parameters α and β are respectively raised to a power representing the number of 'guessing' and 'forgetting' corrections necessary to fit all 'true' pattern vectors to an observed data vector. Similarly $(1-\alpha)$ and $(1-\beta)$ are raised to the number of 'correct' responses in each case. The product of these over all possible response patterns is equation (ii). Multiplying this by the probability that the j^{th} true pattern vector occurs (θ_j), and summing this for all true pattern vectors yields equation (i), which represents Dayton and Macready's general probabilistic model.

Dayton and Macready discuss the use of their model with respect to three cases. In Case A, α and β are unrestricted. That is, there are separate guessing and forgetting parameters for each task, although the same parameters characterize a given task across different 'true' pattern vectors. Unfortunately Case A has been solved to date only for concept attainment models, defined by the authors as models in which each subject responds completely correctly or completely incorrectly to a given set of tasks. While of potential use to those interested in criterion-referenced testing, at present Case A is therefore not of general use in validating learning hierarchies. In Case B all α's are the same for different tasks and all β's are the same for different tasks. In Case C, which is equivalent to Proctor's model, only one misclassification parameter is included. That is, α equals β. In all three cases maximum likelihood estimates of the various α's, β's, and θ's are obtained through a series of iterations. These values are used to compute the number of expected responses for each possible response pattern. The goodness of fit between data and hierarchy model is then calculated by both a Pearson chi-square test and a likelihood ratio expressed in the form of a chi-square. The latter appears to be more useful as it is less severely distorted by small frequencies, an important advantage when it is realized that for all except very small hierarchies a large proportion of the expected frequencies should be nearly zero if the hierarchy is valid.

It appears fair to conclude that the Proctor model, and hence Dayton and Macready's Case C, represents a major gain over the earlier scaling methods. Further, since the precision of Dayton and Macready's general probabilistic model is dependent upon the precision of the misclassification parameter(s), it is clear that their Case B represents an advance over Case C. In theory, Case A, in which the misclassification parameters are unrestricted, represents the most appropriate situation. Since Case A has so far been solved only for the relatively simple concept-attainment model, at the present time the most appropriate application of Dayton and Macready's model, and of scaling in general, to the determination of learning hierarchies is Case B. To date, there do not appear to be any reports of its use in applied situations.

A Comparison of the Advantages and Disadvantages of the Methods Reviewed

In the present chapter the major methods used to validate learning hierarchies have been described. These methods may be sub-divided into those involving scaling, those involving comparisons of the elements in the hierarchy in pairs, and those involving transfer of learning.

Of the first two kinds of methods those involving scaling may be considered conceptually the most pleasing because they consider the hierarchy as a whole--or at least in larger pieces than pairs of elements. These methods, despite severe limitations initially, have progressed to the stage where the most recent development (Dayton and Macready, 1976) offers much promise. It offers a maximum likelihood procedure to test goodness of fit between model and data and is applicable to hierarchies of any configuration. However, it has several important disadvantages. Firstly, the computer programme which is essential to its application can only accommodate small hierarchies at present. Secondly, incorrect response patterns are accommodated to the hierarchical model by means of a guessing parameter common to all elements in the hierarchy and a forgetting parameter similarly common to all elements in the hierarchy. The estimated values of these two parameters affect the predicted frequencies of *all* response patterns, thereby diminishing the potential precision of the model.

Of the models involving pairwise comparisons, that of White and Clark (1973) is easily the most sophisticated. The present investigator feels that the Linke (1976) modification of the test is potentially an improvement, although its implications need more investigation. Dayton and Macready claim that their model subsumes that of White and Clark and that the latter is equivalent to their Case A. Although true in theory this claim is misleading in practice because Dayton and Macready have solved Case A only for the concept attainment model. This suggests, measurement errors excepted, that for each separate skill in the pair under test any subject should answer all questions correctly or incorrectly. This is not implied by White and Clark and is less likely to be true as the number of questions increases. The problems caused by the use of common misclassification parameters in Case B and Case C of the Dayton and Macready model is avoided in the White and Clark test because the latter has effectively a guessing and a forgetting parameter for each skill. In this sense, rather than being subsumed by the Dayton and Macready model the White and Clark test has an advantage. Finally, Dayton and Macready correctly note that the White and Clark test is limited to equal numbers of questions per skill and to no more than three questions per skill. However, for the purpose of the present study the White and Clark test has been generalized to overcome these limitations (Griffiths and Cornish, 1978). The development of this generalization is described in the next section of the present chapter.

The ordering-theoretic method is conceptually less pleasing than the White and Clark test because unlike the latter it does not have a non-arbitrary test of the validity of the hierarchical connections being tested. However, it is much simpler to use and is currently popular in the research literature.

Neither the White and Clark test nor the ordering-theoretic method considers an hierarchy as a whole. It is possible that combining the results of analyzing skills in pairs may lead to a different hierarchy than when the hierarchy is considered as one unit. Moreover, it seems reasonable to assume that validation of an hierarchy as a whole is a more acceptable procedure because in subsequent applications the whole hierarchy is more likely to be used. The Dayton and Macready scaling model allows testing of complete hierarchies of any configuration, although further development to allow for larger hierarchies and unrestricted misclassification parameters is indicated.

In the present study, the White and Clark test, the ordering-theoretic method and the Dayton and Macready method are applied to the same data, offering the potential for a comparison between them as well as increasing the evidence upon which the validity of the hypothesized or alternate hierarchy may be judged. The other psychometric methods reviewed are all considered to be of limited usefulness for hierarchy validation for the reasons already indicated, and will not be applied to the data of the present study.

As has already been indicated it is possible to consider the validity of an hierarchy in terms of whether learning subordinate skills promotes learning of related superordinate skills in the hierarchy. Unfortunately, simple indices derived to test this, such as Gagné's proportion positive transfer, have proven fallible. From the viewpoint of the researcher, if not the student, the most satisfactory means of testing for positive transfer appears to be direct comparison of randomly assigned groups of students taught by following the hierarchy with a similar group taught without the use of the hierarchy, or even taught by a deliberately scrambled hierarchy. Alternatively, and more ethically perhaps, a group needing and given remediation in accordance with the hierarchy might be compared with a similar group which has not received remediation.

In a number of studies either the psychometric or the transfer definition of hierarchical dependency has been applied, but not both. It is here suggested that both aspects are of sufficient importance that an hierarchy validated by either but not both approaches should be regarded as incompletely validated. The fact that it can be shown empirically that one skill (say B) is not learned without prior learning of another skill (say A) does not necessarily mean that learning A helps a group of individuals to learn B. Conversely, a significant positive correlation between the learning of the two skills does not mean that the learner must master A first. However, if it can be reliably demonstrated that B cannot be learned until A is learned, *and* that learning B is associated with prior learning of A, then it can be claimed more legitimately that the skills are in hierarchical relationship to one another. This is the view taken in the present study.

Extension and Modification of the White and Clark Test

The limitation of the White and Clark test to equal numbers of questions per skill and to no more than three questions per skill restricts its usefulness. To overcome this restriction Griffiths and Cornish (1978) have generalized the test to allow for any number of questions per skill and have removed the restriction of equal numbers of questions per skill. This modified version of the White and Clark test is the one used in the present study. The modification is described below.

Consider again the problem of testing for hierarchical connections between pairs of skills. Suppose now that subjects are required to answer m questions on the hypothesized lower skill and n questions on the upper skill, when $m,n \geq 2$. Using White and Clark's notation it is a simple matter to show that the probability Pij under the null hypothesis of a subject answering exactly i questions correctly on the lower skill and exactly j questions correctly on the upper skill is given by

$$(1) \quad P_{ij} = \binom{m}{i}\binom{n}{j} \left(P_o \theta_b^{\,i}(1-\theta_b)^{m-i}\theta_d^{\,j}(1-\theta_d)^{n-j} + P_I \theta_a^{\,i}(1-\theta_a)^{m-i}\theta_d^{\,j}(1-\theta_d)^{n-j} + P_{II}\theta_b^{\,i}(1-\theta_b)^{m-i}\theta_c^{\,j}(1-\theta_c)^{n-j} + P_B\theta_a^{\,i}(1-\theta_a)^{m-i}\theta_c^{\,j}(1-\theta_c)^{n-j}\right)$$

$i = 0,1, \ldots, m$; $j = 0,1, \ldots, n$. Summing (1) with respect to j, the expected number of subjects answering exactly i questions correctly on the lower skill is

$$(2) \quad N\sum_{j=o}^{n} P_{ij} = N\binom{m}{i}\{Q\theta_a^{\,i}(1-\theta_a)^{m-i} + (1-Q)\theta_b^{\,i}(1-\theta_b)^{m-i}\}$$

$i = 0,1, \ldots, m$, where N = total number of subjects in the sample, and $Q = P_I + P_B$. Similarly, summing (2) with respect to i, the expected number of subjects answering exactly j questions correctly on the upper skill is

$$(3) \quad N\sum_{i=o}^{n} P_{ij} = N\binom{n}{j}\{R\theta_c^{\,j}(1-\theta_c)^{n-j} + (1-R)\theta_d^{\,j}(1-\theta_d)^{n-j}\}$$

$j = 0,1, \ldots, n$, where $R = P_{II} + P_B$. We denote by a_i, $i = 0,1, \ldots, m$, the observed number of subjects answering exactly i questions correctly on the lower skill and by b_j, $j = 0, 1, \ldots, n$, the number of subjects answering exactly j questions correctly on the upper skill.

Following a similar procedure to that of White and Clark (1973) we can obtain estimates $\hat{\theta}_a$, $\hat{\theta}_b$, $\hat{\theta}_c$, $\hat{\theta}_d$ of the parameters θ_a, θ_b, θ_c, θ_d. Equating, for example, the expected conditional probabilities of a subject answering m questions on the lower skill correctly given that he has m-1 or [illegible] obtained from (2), (3), to the estimate obtained from the obse[illegible], we have

$$\frac{a_m}{a_{m-1} + a_m} = \frac{Q\theta_a^{\ m} + (1-Q)\theta_b^{\ m}}{m[Q\theta_a^{\ m-1}(1-\theta_a) + (1-Q)\theta_b^{\ m-1}(1-\theta_b)] + [Q\theta_a^{\ m} + (1-Q)\theta_b^{\ m}]},$$

$m \geq 3$.

Using the prior knowledge that $1-\theta_a$, θ_b, $1-\theta_c$, θ_d will be close to zero and ignoring terms in powers of $1-\theta_a$, θ_b, $1-\theta_c$, θ_d higher than the first we obtain, after rearrangement, the estimate

$$(4) \quad \hat{\theta}_a = \frac{ma_m}{ma_m + a_{m-1}}, \quad m \geq 3.$$

Similar arguments yield

$$(5) \quad \hat{\theta}_b = \frac{a_1}{a_1 + ma_0}, \quad m \geq 3.$$

$$(6) \quad \hat{\theta}_c = \frac{nb_n}{nb_n + b_{n-1}}, \quad n \geq 3.$$

$$(7) \quad \hat{\theta}_d = \frac{b_1}{b_1 + nb_0}, \quad n \geq 3.$$

We denote by $\bar{x}$, $\bar{y}$ the observed average numbers

$$\bar{x} = \frac{1}{N} \sum_{i=0}^{m} ia_i \quad , \quad \bar{y} = \frac{1}{N} \sum_{j=0}^{n} jb_j$$

of correct answers per subject on the lower and upper skills respectively. The corresponding expected values obtained from (2), (3) are, after simplification,

$$\frac{1}{N} \sum_{i=0}^{m} i \{N \sum_{j=0}^{n} Pij\} = m [Q\theta_a + (1-Q)\theta_b]$$

and $\frac{1}{N} \sum_{j=0}^{n} j \{N \sum_{i=0}^{m} Pij\} = n [R\theta_c + (1-R)\theta_d]$

Equating observed and expected average numbers and using (4) - (7) and the above approximations we obtain, after rearrangement, the estimates

$$(8) \quad \hat{Q} = \frac{\bar{x}/m - \hat{\theta}_b}{\hat{\theta}_a - \hat{\theta}_b}$$

and

$$(9) \quad \hat{R} = \frac{\bar{y}/n - \hat{\theta}_d}{\hat{\theta}_c - \hat{\theta}_d}$$

The estimates of P_I and P_O and P_B are related to $\hat{Q}$, $\hat{R}$ and $\hat{P}_{II}$ by the equations

$$\hat{P}_I = \hat{Q} - \hat{R} + \hat{P}_{II},$$

$$\hat{P}_O = 1 - \hat{Q} - \hat{P}_{II},$$

$$\hat{P}_B = \hat{R} - \hat{P}_{II}.$$

In the case of White and Clark's test, the test of significance of the null hypothesis H_o: $P_{II} = 0$ against the alternative H_a: $P_{II} > 0$ is performed by first calculating an estimate $\hat{P}_{on}$ of P_{on} by substituting the above estimates of the parameters (under H_o) into (1). Denoting the observed frequency in cell (i,j) by f_{ij}, the significance of the observed data is given by the binomial expression

$$\text{(10)} \quad \text{sigdata} = 1 - \sum_{X=0}^{f_{on}-1} \binom{N}{X} \hat{P}_{on}^{\,X} (1 - \hat{P}_{on})^{N-X}$$

If sigdata is less than a prescribed significance level, H_o is rejected. The White and Clark test depends, therefore, upon the observed and expected frequencies in the critical cell (o,n). When m and n are two or three this procedure is quite appropriate. However, when m or n exceeds three there is a case for increasing the number of critical cells. If S denotes the set of critical cells chosen as appropriate by the experimenter then the significance of the observed data is given by the multinomial expression

$$\text{(11)} \quad \text{sigdata} = 1 - \sum \frac{N!}{N_o!} (1-\hat{p})^{N_o} \prod_{(i,j)\varepsilon S} \frac{P_{ij}^{\,x_{ij}}}{x_{ij}!}$$

Where Σ is taken over all combinations of non-negative integers x_{ij} such that

$$o \leq \sum_{(i,j)\varepsilon S} x_{ij} < \sum_{(i,j)\varepsilon S} f_{ij} , \quad \text{and} \quad N_o = N - \sum_{(i,j)\varepsilon S} x_{ij} ,$$

$$\text{and} \quad \hat{p} = \sum_{(i,j)\varepsilon S} P_{ij} .$$

It is readily seen that (11) reduces to (10) in the case of a single critical cell (o,n).

As an illustration, in the case of $m = n = 4$ observations falling in the cells (0,3), (0,4), (1,3), (1,4) might be considered as evidence against H_0. These critical cells are shaded in Figure 5.

In the present study application of the extended and modified White and Clark test involves use of two, four and eight test items per skill. The critical cells believed to be appropriate in these cases are as follows:

$m = 2$, $n = 2$: (0,2)

$m = 2$, $n = 4$: (0,3), (0,4)

$m = 2$, $n = 8$: (0,6), (0,7), (0,8)

$m = 4$, $n = 2$: (0,2), (1,2)

$m = 4$, $n = 4$: (0,3), (0,4), (1,3), (1,4)

$m = 4$, $n = 8$: (0,6), (0,7), (0,8), (1,6), (1,7), (1,8)

$m = 8$, $n = 2$: (0,2), (1,2), (2,2)

$m = 8$, $n = 4$: (0,3), (0,4), (1,3), (1,4), (2,3), (2,4)

$m = 8$, $n = 8$: (0,6), (0,7), (0,8), (1,6), (1,7), (1,8), (2,6), (2,7), (2,8).

In the case where one of m or n equals two the above theory holds almost unchanged. If $m = 2$ and $n > 2$, equations (4) - (9) are unchanged except that (5) must be replaced with $\hat{\theta}_b = 0$. If $m > 2$ and $n = 2$, equations (4) - (9) are again unchanged except that (6) must be replaced with $\hat{\theta}_c = 1$. (The rationale that White and Clark use to put $\hat{\theta}_b = 0$, $\hat{\theta}_c = 1$ in their case of $m = n = 2$ may be suitably extended in the present case.)

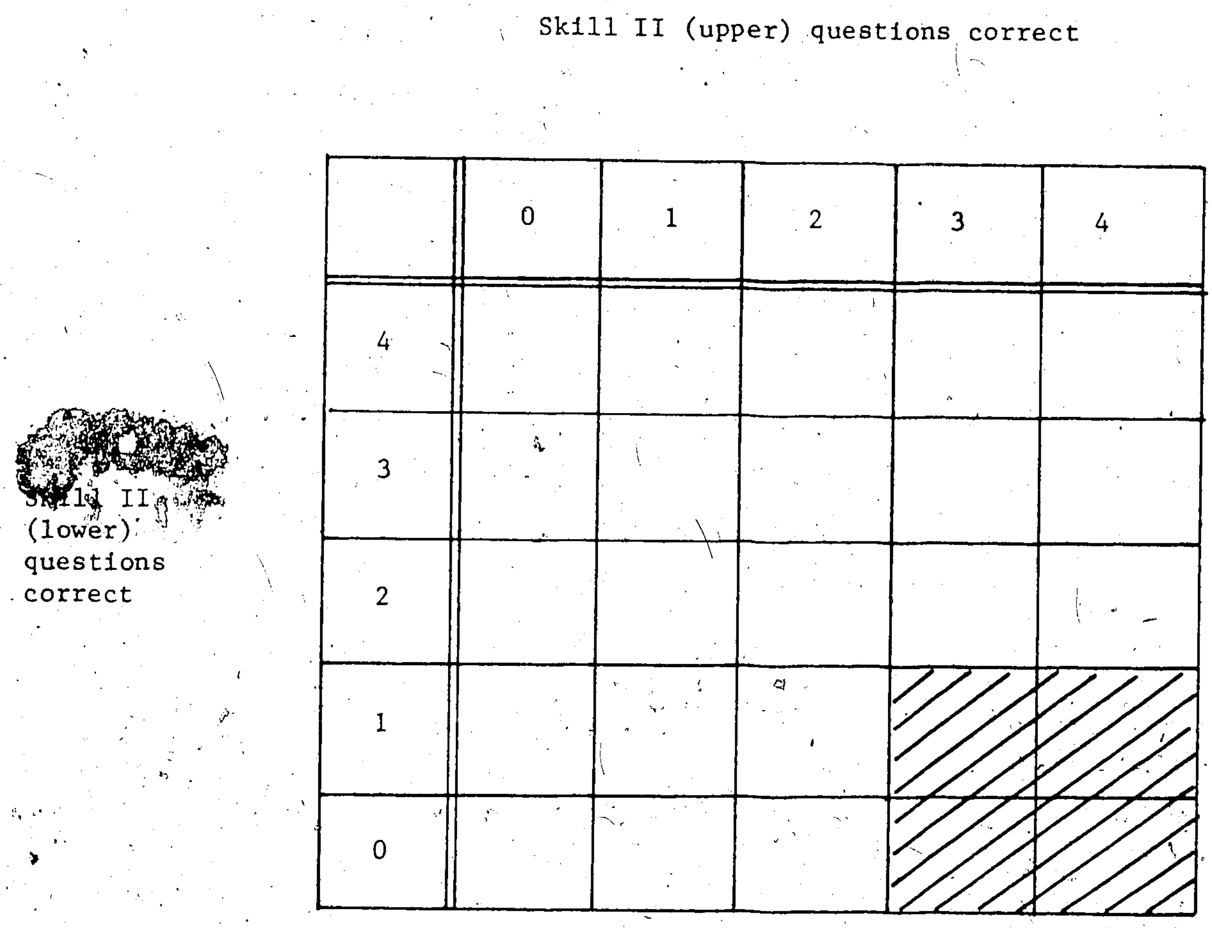


Fig. 5. Skill by skill matrix for modified White and Clark test

Criterion-Referenced Testing

The tests used to determine mastery status of individuals with respect to the intellectual skills in a proposed hierarchy should clearly be criterion-referenced. This chapter concludes with a discussion of the parameters of such tests.

Glaser and Nitko (1971) define a criterion-referenced test as "one that is deliberately constructed to yield measurements that are directly interpretable in terms of specified performance standards." Woodson (1973) notes that criterion-referenced scales have meaning relative to the characteristic measured rather than the distribution of the characteristic in some population. Criterion-referenced measurement is clearly of interest in learning hierarchy research. The last few years have been notable for a consideration of the particular statistical characteristics of criterion-referenced measures.

Gagné (1969a) notes:

> "What is measured" should be the major issue in the development of criterion-referenced measures. Therefore, primary attention should be given to the single item. Characteristics of criterion-referenced measurement are: 1) distinctiveness of items in measuring a particular class of performance, 2) freedom from distortion arising from sources other than learning itself, 3) scoring based on the single item rather than a test, 4) inapplicability of the concept of difficulty, since items should be distinctive and free from distortion, 5) establishment of reliability by use of two items only from a single class of behavior, and 6) appropriateness of "content validity" rather than predictive validity.

There is general agreement with Gagné's emphasis on content validity. Hambleton and Gorth (1971) suggest that "a carefully made judgement based on the test's apparent relevance to the behavior legitimately inferable from those delimited by the criterion" represents the most relevant aspect of validity. Popham and Husek (1969) earlier made the same point and also raised the important issue that the classical statistical tests for reliability and validity depend on variability of test scores, but that for criterion-referenced tests variability is irrelevant. Hence, in their opinion, the classical procedures are not applicable. For example, there is nothing wrong with a criterion-referenced test if, after instruction, everyone obtains a perfect score. However, this would lead to a zero estimate of internal consistency when conventional reliability estimates such as the KR20 formula are applied. Measures of validity are similarly affected.

There is general, although not universal, agreement with this position. Opposition is centred around two concerns. Woodson (1973) argues that items and tests which give no variability give no information and are therefore not useful, and further that facility should be much higher after instruction than before and this difference may be used to provide the necessary variability in scores. The same direction was also pursued by Haladyna (1974) and Wilson and Burnett (1975). A second school of thought suggests that it would be better to modify existing statistical procedures to fit criterion-referenced testing than to reject out of hand hard-won existing procedures. The major proponent of this view has been Livingston. He suggests (1972) substitution of deviation from the mastery criterion score instead of deviation from the mean as in conventional reliability tests. Such a position has been criticized by Hambleton and Novick (1973) who contend that a deviation score is not relevant to mastery learning, by Harris (1972) on technical grounds and by Oakland (1972) who notes that tests with few items (typical in criterion-referenced testing) could not be used. The general consensus is that statistical tests for criterion-referenced measurement must be specifically developed.

After dismissing as inappropriate those tests which depend upon variance of test scores Brennan (1974), in a comprehensive analysis, concludes that the appropriate measure of reliability is internal consistency in terms of test-retest stability. He further concludes that where a mastery cutting score is involved, as is most often true in criterion-referenced testing, an index suggested by Hambleton and Novick (1973) appears to have the most appealing theoretical rationale. Hambleton and Novick's index assumes two administrations of the same test or parallel forms of the same test. Assuming examinees are to be classified into m mastery states, the index of reliability P_o is represented by $P_o = \sum_{k=1}^{m} P_{kk}$ where P_{kk} is the proportion of examinees classified in the kth mastery state on the two administrations.

Swamimathan, Hambleton and Algina (1974) note that Hambleton and Novick's index does not allow for the proportion of agreement due to chance alone and may therefore yield misleading estimates of reliability. To take account of this problem they suggest use of a coefficient (κ) developed by Cohen (1960) as an index of reliability. This coefficient is defined as $\kappa = (P_o - P_c)/(1-P_c)$ where

$$P_c = \sum_{k=1}^{m} P_{k\cdot}P_{\cdot k}$$

and $P_{k\cdot}$ and $P_{\cdot k}$ represent the proportions of examinees assigned to mastery state k on the first and second administrations, respectively, and P_c represents the proportion of agreement that would occur even if the classifications based on the two administrations were statistically independent. The characteristics and uses of κ have also been exemplified by Light (1973). Cohen's coefficient is used to estimate the reliability of the chemistry tests used in the present study.

Returning to the question of validity, Hambleton and his colleagues (1975) reiterated the central role of content validity, and also noted the general agreement that an efficient way to ensure this is to carefully specify the domain to be tested and then generate a bank of items corresponding closely to the domain specifications. Such item generation typically follows the item-form procedure developed by Hively et al. (1968). It may be suggested, however, that more narrowly specified objectives may be perfectly well represented by a smaller number of items and that little is to be gained by specifying a larger number of virtually identical items. This appears to be Gagné's position and is the one taken in the proposed study. However the test items are generated, expert agreement of the content validity of the test is vital. In the paper by Hambleton and his colleagues a procedure is described for measuring such agreement, making allowance for chance agreement.

The number of questions necessary to reliably measure subject mastery of a particular objective is clearly of concern. Roudabush and Green (1972) suggest that more specific objectives require less test items. They found that three or four items per objective were sufficient in such a situation.

Summary

This chapter has reviewed the methods which have been used to validate learning hierarchies, including some recent extensions. One of these methods, the White and Clark test, has been extended for use in the present study. The chapter concludes with a consideration of the parameters of criterion-referenced tests as these appear to be a necessary component to the determination of the mastery status of individuals with respect to the skills of which learning hierarchies are composed.

The next chapter is concerned with the application of some of the methods described in this chapter to the validation of learning hierarchies in science.

Chapter 3

HIERARCHIES IN SCIENCE

Learning Hierarchies

Gagné's early empirical studies relating to learning hierarchies were all concerned with the development of mathematics skills. The majority of published studies relating to learning hierarchies since that time have been concerned with mathematics and science, with the emphasis much more towards mathematics. Indeed, few well established learning hierarchies exist in the area of science, and these often contain a substantial proportion of mathematics skills. The present chapter is therefore brief but represents the extent of the published literature relating to learning hierarchies in science. The chapter begins with a reference to an elementary science curriculum 'Science A Process Approach' (SAPA). This is followed by a discussion of empirical studies and comments relating to learning hierarchies in chemistry and physics. Finally, several reports of hierarchies based upon Piagetian theory will be discussed.

The best known and most extensive attempt to apply Gagné's hierarchical model in science is 'Science A Process Approach' (1965), a curriculum for children from kindergarten to grade six. In this curriculum the focus of instruction is different to what is found in the other studies reported in this chapter, in that the skills to be developed represent what the authors considered to be important operational processes involved in doing science, rather than the development of concepts. The processes observing, classifying, measuring, communicating, inferring, predicting, using space-time relationships, and using numbers were considered to be basic and were given greater emphasis in the early years. Five other processes, namely controlling variables, formulating hypotheses, interpreting data, defining operationally and experimenting were considered to be integrated processes to be developed after reasonable facility with the basic processes had been attained. For each process an hierarchy was constructed leading the learner to what was considered by the authors to be an acceptable level of proficiency with regard to the particular process. The result was an integrated network of hundreds of skills. Reference has already been made in Chapter 1 to the fact that Gagné now recommends that the amount of content to be represented in an hierarchy should be very much smaller than that represented in SAPA. Certainly the cumulative level of success exhibited by students enrolled in the SAPA experimental classes was less than the authors hoped. One reason for this may have been the use of Walbesser's indices for validation of the hierarchical relationships involved, given the limitations already discussed for these indices. Another possible reason may be inferred from a comment by Gagné (1973, p. 25) to the effect that the SAPA learning hierarchy is not a learning hierarchy at all because the instructional units involved are entire lessons and the contents of individual lessons were not designed as learning hierarchies.

The usefulness of the learning hierarchy model to individualized instruction is noted from time to time. For example DeRose (1969), using the 'Chemical Bond Approach' (CBA) materials developed, but did not validate, an hierarchy of 86 basic and 82 optional objectives. Similarly, Boblick (1971) indicates the potential usefulness of learning hierarchies in the application of the systems approach to curriculum development in science. In his report an hierarchy leading to writing chemical formulae as the terminal skill is indicated, but no mention is made of its validation.

A carefully designed study by Okey and Gagné (1970), described in another communication in greater detail (Okey and Gagné, 1971), involved use of a programmed unit on solubility product. The unit was first analyzed by the investigators to identify the implicit hierarchy involved. Two equivalent tests were constructed for the superordinate skill. In addition two equivalent tests were constructed for the collective subordinate skills. The sample consisted of 135 tenth, eleventh, and twelfth grade students in five chemistry classes. Two equal groups within each class were randomly formed. All students were administered a pre- and post-test on what the authors had hypothesized to be the necessary subordinate skills. The first group completed the unit and testing, while the second group did laboratory work unrelated to the unit. The unit was revised by adding further instruction on subordinate skills failed by many students, and then administered to the second group. The result was a significantly better performance for the second group compared to the first (significant at the .001 level). The authors appropriately concluded that this result supports the cumulative learning model. In particular, they noted that the effect of learning the subordinate skill on learning of the superordinate skill was cumulative in the sense that subordinate skills later in the sequence exert a disproportionate influence.

Despite the attractiveness of Okey and Gagné's study several critical comments are in order. First, the skills involved in the study are not described as precisely as they might be. In particular it may be noted that the description of the superordinate skill is to "solve solubility product problems," which could clearly encompass a wide range of outcomes. Secondly, for each of nine out of fifteen subordinate skills less than 80 percent of the experimental group were successful. For four of these skills less than 40 percent were successful. The effect of lack of these subordinate skills for individual subjects was not investigated, nor were specific transfer effects between skills. It is quite possible therefore that the hierarchy involved may be less valid in terms of both its psychometric and transfer characteristics than the encouraging results suggest.

A study by Seddon (1974) was concerned with the development of students' understanding of the 'Kimball Charge Cloud Model' of chemical bonding. The sample, between 15 and 19 years of age, consisted of 533 students in schools preparing for 'O' and 'A' level chemistry examinations and 208 students enrolled in first year university or teacher training college chemistry courses. All subjects were instructed through a self-instructional unit prepared by the author. Before commencing this unit a pre-test based on the content of the unit was administered. At the end of instruction the same test was used as a post-test. The purpose of the study was to determine the relative effectiveness of the pre-test, a general chemistry test administered before commencement of the study, intelligence as measured by a standardized intelligence test, and age as predictors of achievement on the post-test. A regression analysis indicated that general chemistry knowledge was the best predictor followed closely by the pre-test. The author interpreted this as supporting Gagné's hierarchical model, in that both the general chemistry test and the pre-test were contributing a factor which might be labelled related knowledge. However, this support may be misplaced because the general chemistry test in particular was not concerned with capabilities specifically prerequisite to skills tested in the post-test. It is here suggested that the results may reflect no more than a general disposition towards chemistry, ability in chemistry or some other such factor.

A study reported recently by Gower, Daniels and Lloyd (1977) is particularly relevant to the present study as it is also concerned with identification of an hierarchy concerning the mole concept. The superordinate element in their hierarchy represents understanding of the concept of molar mass. However the authors comment that their initial theoretical analysis indicated two independent hierarchies, one consisting of concepts based on empirical experience and the other representing an hierarchy of theoretical concepts. As in the present study, the data for the hierarchical analysis were obtained by requiring the sample (N = 42) to respond to a set of test items representing the elements of the hypothesized hierarchies. The results of the top 27 percent and the bottom 27 percent of the sample were used for analysis. Each element in each hierarchy was tested by four items representing recall, comprehension, application and analysis, respectively. The possibility of an hierarchical relationship between each element and each of those hypothesized to be subordinate or equivalent to it was tested by applying a consistency ratio defined as

$$\text{Consistency} = \frac{A + B + C}{A + B + C + D}$$

where the frequencies A, B, C, D are defined as for Figure 2 (p. 46) but with only one lower skill in each case. To determine the values of these frequencies part values were first obtained by comparing the responses to similar items (for example recall with recall) for each pair of skills. The part values were then totalled to give the frequencies A, B, C and D. A value of 0.85 for the consistency ratio was arbitrarily considered acceptable evidence for the existence of an hierarchical connection between two skills. The authors claimed that their results supported the hierarchical model. However, examination of their data suggests otherwise. For the empirical hierarchy only 12 out of 18 connections in the 'validated' hierarchy showed a consistency ratio of 0.85 or more. For the theoretical hierarchy the results were much worse. In this case only seven out of 22 connections in the 'validated' hierarchy achieved the critical value for the consistency ratio. It appears that the results deny rather than support the existence of an hierarchy leading to the mole concept. However, in the opinion of the present investigator the conduct of the study by Gower, Daniels and Lloyd does not allow firm conclusions in either direction for the following reasons. First, the consistency ratio is not an appropriate measure of hierarchical dependency for the reasons already discussed with respect to the use of Gagné's and Walbesser's indices. Indeed, the consistency ratio may be a worse measure than Gagné's proportion positive transfer because its value will always be higher than proportion positive transfer which itself is biased towards acceptance of connections being tested. When the consistency ratio is applied any connection being tested will be accepted if less than 15 percent exceptions are found, an unusually liberal criterion. A second objection relates to the kinds of capabilities actually tested by the test items. Although the authors refer to the testing of intellectual skills, their understanding of this term is not consistent with Gagné. The items testing recall represent verbalized knowledge, those testing analysis are likely to represent the use of cognitive strategies and some of the items testing at the application level may also represent the use of cognitive strategies, depending on the unfamiliarity of the context. It seems likely that only the items at the comprehension level are consistent with Gagné's understanding of what is meant by intellectual skills, and these represent only one item testing a particular element. The results do not therefore appear to be interpretable in terms of the possible existence of an hierarchy for the mole concept.

Developmental Hierarchies

In a number of studies hierarchies have been derived from naturalistic descriptions of intellectual development reported by Piaget and others rather than by a logical task analysis procedure. Such hierarchies are described as developmental hierarchies. Some of these are important to the present study in that they describe the development of scientific concepts.

Resnick (1973) notes "that stage theories of development, such as Piaget's, are hierarchical in that they propose that individuals can reach a higher stage of development only by passing through a fixed series of lower stages." Resnick also notes that invariant sequences of cognitive development have also been proposed for much smaller units of behavior. It is this level of developmental hierarchy that is of particular interest in the present discussion. However, despite similarities in terms of structure and methods of validation there appear to be fundamental differences between learning hierarchies and developmental hierarchies.

Learning hierarchies are usually derived by task analysis focusing upon the characteristics of skilled performance of the task and its prerequisites. Identification of unskilled performance forms no part of this procedure. Conversely, developmental hierarchies are likely to be concerned with unskilled as well as skilled performance as they appear in the natural setting. The difference may be critical. According to Resnick (1973) "Hierarchies generated in this way will tend to treat errors and misunderstandings typical of earlier stages as <u>necessary</u> prerequisites for the individual to reach higher stages." In the developmental model the resolution of these difficulties by the learner himself is necessary to the development of his intellectual structures. In this way 'equilibration' occurs. Shulman (1968) distinguishes this from the Gagnéan model as a roller-coaster ride of successive equilibria and disequilibria versus a smoothly guided tour up a carefully constructed hierarchy of objectives. Attempts to smoothly induce specific vertical transfer through a learning hierarchy may therefore be antithetical to the developmental hierarchy model. It is not surprising then that developmental hierarchies are usually validated by psychometric (usually scaling) methods rather than by the determination of significant positive transfer. In summary, the learning hierarchy is prescriptive whereas the developmental hierarchy is descriptive. Despite these differences it seems likely that research into either model may be of benefit to those concerned to articulate the other. Indeed such a combination may be a particularly profitable direction for research. The studies reflecting developmental hierarchies which will now be reviewed are therefore considered to be of potential use to those concerned with learning hierarchies in science. Several studies which relate the Gagnéan and Piagetian models are of particular interest to the present study.

Kofsky (1966) attempted to verify Piaget's description of the order in which children acquire classification skills. The responses of 122 children to a series of eleven tasks ranging from resemblance sorting to hierarchical classification were analyzed. The sample ranged from four to nine years of age, and the tasks were administered by individual interviews. The results indicated that the subjects varied in the sequence in which they were able to successfully solve the tasks. The value of a modified Guttman reproducibility coefficient for the hypothesized scale was too low to merit acceptance of the scale as valid. The results were therefore contrary to Piaget's suggested order of development of classification skills. Other possible interpretations suggested by Kofsky include difficulties of ensuring consistent task administration, irregular performance by young children and finally that scalogram analysis represents an abrupt picture of development at the particular time and may not be useful when considering a model of continuous development.

The development of classification skills is of fundamental importance in the learning of science. It is appropriate therefore that Kofsky's study should be replicated. This was undertaken by Allen (1970) who presented Kofsky's tasks to 190 elementary school children. Again the eleven items did not form a unidimensional scale. However the application of Lingoes' modification of Guttman scaling yielded several smaller scales and suggested at least a partial ordering with grouping skills emerging first and class inclusion skills last. Like Kofsky, Allen preferred to consider other interpretations of the results in addition to the failure of the skills involved to form a scale. In particular Allen noted the potential unreliability of testing the individual classification skills by one item each.

Raven (1968) examined the development of the concept of momentum in children between five and eight years of age, and in doing so compared the appropriateness of a developmental hierarchy and a learning hierarchy as models of the development of the concept of momentum. According to the developmental hierarchy, derived by Raven from Piaget's writings, the child acquires the concept of momentum followed in order by conservation of matter, proportional use of mass and speed with momentum held constant and finally the concept of speed.

According to the learning hierarchy derived by logical analysis by Raven the expected order of acquisition is conservation of matter, speed, proportional use of mass and speed with momentum held constant and finally acquisition of the concept of momentum. To test the appropriateness of these two alternative hierarchies 160 children were selected randomly and individually administered a set of six tasks representing the concepts involved in the hierarchies. The order of administration of the tasks was randomized. An interview format was applied for administration of the tasks. The tasks were compared according to their level of observed difficulty. The results favored acquisition of the concepts involved in the order represented by the developmental hierarchy. However, this interpretation may be less certain than the author suggests for several reasons. Firstly, once again only one task was used to test each concept, with the exception of the concept of momentum where two tasks were used. Secondly, the task testing understanding of speed was perceptually different to the other tasks in that the subjects could not directly observe the objects whose speed was being compared. This task may therefore be spuriously difficult. A more important criticism is that the hypothesized learning hierarchy may not be a learning hierarchy at all. It seems likely that this hierarchy was derived by re-combining the components of the developmental hierarchy in a 'logical' order. The steps involved in the hierarchy are very large and it may be reasonably suggested that a more precisely defined hierarchy may yield different results. Finally, the level of understanding of the concepts involved may be of importance. In a learning hierarchy mastery of the component skills is required if further progress is to be made through the hierarchy. In the study under discussion understanding of the concept of momentum is considered only to an intuitive level. It is therefore not surprising that a logical hierarchy is not substantiated by the data.

A later study by Raven (1972) tested Piaget's description of the child's development of the concept of acceleration. Twenty-four children from each of the grades three to six were selected randomly and administered seven tasks relating to the concepts speed and acceleration. Each task was administered in an interview format to individual subjects. Application of scalogram analysis to the data suggested moderate support for a scale reflecting Piaget's proposed order of development, with a reproducibility of 0.86 being observed. The scale reflects acquisition of intensive concepts requiring comparison of the whole with part before acquisition of extensive concepts which require comparison of parts with one another. The author concluded that ". . . simultaneous and successive motion activities involving the use of intensive operations could be presented to third and fourth grade children. The results also suggest that activities involving successive motion that uses extensive logical operations could be used with some fifth and sixth grade children."

Phillips (1971) describes another study which is similar to those of Raven in that the intent is to attempt to use Piaget's writings to develop a fine structure for the development of selected science concepts by children. Phillips derived an hierarchy consisting of twelve concepts leading to the conservation of displaced volume. Phillips selected six of these concepts and these were tested in four tasks. The tasks were administered to a sample consisting of 40 randomly selected children in each of grades three, five, and seven. Two modes of presentation were used, one in which the objects involved in the tasks were present and the other in which graphic representations of the same tasks were used. Each mode was applied to only half the sample, and required oral presentation of the tasks as well as oral responses. The data for each group were analyzed separately to determine if the hypothesized hierarchy formed a Guttman scale. In each case the reproducibility did not exceed the prespecified criterion of 0.90. Hence the hierarchy was rejected. However, the order observed was the same for both modes of presentation and in each case only one element deviated from the hypothesized order. Phillips correctly notes that the deviance of one item could be explained by a faulty task as well as by a faulty hierarchy, and suggests that the sequence may be considered useful in teaching conservation of displaced volume. It may be suggested, however, that the selection of only six of the twelve concepts from the original hierarchy does not allow an interpretation of the validity of the complete hierarchy.

Each of the developmental hierarchy studies reported above was concerned with verifying Piaget's descriptions of the development of particular concepts. In no case was instructional intervention employed to determine the transfer properties of these hierarchies. A study reported by Bass and Montague (1972) involved deriving two hypothesized hierarchies from Piaget's description of children's responses to the 'balance' task and the 'inclined plane' task (Inhelder and Piaget, 1958), and then developing programmed units of instruction leading to performance of each task. A pre-test, instruction and post-test sequence on each of these units was applied to 133 ninth grade science students, most of whom were 14 or 15 years of age. The validity of the two hierarchies was evaluated on the basis of the percentage of correct responses to items testing each level of the hierarchy. The authors correctly note that analysis of group responses rather than the pattern of responses of each individual represents a necessary but not a sufficient criterion of the validity of an hierarchy. The pre-test and post-test data supported the existence of the hierarchy leading to the application of a quantitative rule for equilibrium in the balance. The proportion of subjects able to exhibit this behavior increased from 45 percent before instruction to 75 percent after instruction. The authors interpreted this to mean that the instructional treatment was effective.

This result appears to offer further support to the hierarchy because it appears to induce positive transfer. However, it should also be noted that the overall performance of the sample was already very high on the pre-test for the four prerequisite levels, and showed little or no increase after instruction. In particular 82 percent succeeded at the penultimate level on the pre-test while 84 percent succeeded at this level on the post-test. It seems quite possible therefore that the improved performance at the ultimate level of the hierarchy may be ascribed to the effect of the pre-test rather than instruction, especially as the total instructional time was only three hours. In summary, the data does not deny the existence of the 'balance' hierarchy, but neither are they very supportive. In the case of the second hypothesized hierarchy leading to the application of a quantitative relationship enabling prediction of the movement of a small toy wagon on an inclined plane in terms of the weight of the wagon, inclination of the plane and the weight of a counterweight, the hierarchy was not supported by the pre-test or post-test data. Not surprisingly the instructional sequence based on this invalid hierarchy was not very successful.

A study by Wiegand (1969) used a variation of the inclined plane task as the 'Final Task' in an hypothesized hierarchy and the inclined plane task as a test of transfer from lower skills to the Final Task. The Final Task itself involved deriving the relationship between the height of a car on an inclined plane, the weight of a block at the bottom of the plane and the distance travelled by the block after the car was allowed to run down the plane and strike it. The hierarchy hypothesized by Wiegand was very different in structure to the hierarchy derived by Bass and Montague. The latter reflects the psychological development of individuals as suggested by Piaget's naturalistic descriptions, whereas the former represents a sequence of hypothetically prerequisite intellectual skills. The difference is important theoretically as it offers a direct comparison of the use of the Gagnéan and Piagetian models. Wiegand's hierarchy was much more detailed than that of Bass and Montague and the two hierarchies also differed in that in Wiegand's hierarchy the skills were mostly mathematical. Wiegand's study involved 30 randomly selected subjects who were divided into three equal groups. The subjects in each group were individually administered the Final Task and Transfer Task as a pre-test and again as a post-test. None of them could answer the tasks correctly in the pre-test. The members of one group, labelled 'Demonstration Test Retest (DTR),' were shown a demonstration of one example of a test question for each of the sub-skills and were then administered two or more test questions for that sub-skill. This procedure was repeated for each skill downwards through the hierarchy until correct responses were made on tests of two contiguous related sub-skills or until all problems had been administered. Each subject was then retested upwards on each skill not exhibited in the pre-test, but without further demonstration of examples. The post-test on the Final Test and Transfer Test were then administered. The second group, labelled Test Retest (TR) were treated similarly except that no demonstration of examples was given. The third group, labelled Test (T), were administered the initial test for the sub-skills but no demonstration and no retesting of these skills was administered. There were very few exceptions to the hierarchy in any of the three groups.

The DTR group did not differ significantly from the other two groups on the number of initial tests passed for the sub-skills, indicating that demonstration had no effect. In addition no subject passed the post-test on the Final Task without also passing either the initial test or re-test for the subordinate skills. Finally the transfer effect of learning the subordinate skills was shown by the fact that 19 out of 20 subjects in the DTR and TR groups were able to respond correctly to the Final Task and Transfer Task having attained the needed sub-skills, whereas in the T group only three out of ten subjects were able to respond correctly to the Final Task and Transfer Task in the post-test, and each of these subjects passed the initial test for the skill immediately subordinate to the Final Task. Carroll's (1973) suggestion that Wiegand's study demonstrates the effectiveness of immediate experience of component skills rather than that learning of these skills is prerequisite to learning the superordinate task seems a harsh judgment in view of the relative lack of gain on Final and Transfer tasks by the T group. The present investigator prefers to share Wiegand's interpretation that the results are supportive of the need to learn or recall related subordinate intellectual skills when attempting to solve a new task, and hence that the data supports the Gagnéan model of learning rather than the Piagetian model of intellectual development.

Summary

A distinction between learning hierarchies and developmental hierarchies has been made. Despite this there is support for the application of each model to the other, perhaps particularly from the developmental model to the learning hierarchy model. The number of studies reporting well-validated learning and developmental hierarchies is relatively small. However, these studies offer sufficient support to warrant further research into the application of each model.

Chapter 4

DEVELOPMENTAL LEVEL AND ACHIEVEMENT IN SCIENCE

A substantial proportion of the science education literature of the last decade has been concerned with the application of Piaget's theories of intellectual development to science curriculum and instruction. In the present chapter several aspects of this literature which are of direct relevance to the present study are discussed. In the order in which they are presented these include reference to the developmental levels of high school science students, the relationship between learner developmental level and achievement in science, and finally the use of group testing procedures to measure individual intellectual developmental levels.

The Developmental Level of High School Students

In summary of the research to that time Howe (1974) noted that the majority of high school students do not appear to exhibit substantial formal operational thought. Chiappeta (1976) concurred with Howe's finding after reviewing a number of other studies all of which involved administration of at least three developmental tasks to individuals in an interview setting. Indeed, several of the studies reviewed by Chiappeta refer to college students but the same low level of intellectual development was reported. Percentages of high school subjects reported to be formal operational varied between 14 percent (Renner and Stafford, 1972) to 78 percent for a group of chemistry students (Lawson, 1974). Typically less than half of the subjects involved in the various studies reported by Howe and Chiappeta were judged to be formal operational from the responses exhibited. Similarly, Renner (1977) found that of a large sample of subjects in grades ten to twelve in the United States 70 percent failed to exhibit formal operational thinking on the chemical combinations task (Inhelder and Piaget, 1958), 68 percent did not exhibit possession of the proportionality schema as evidenced by the balance task (Inhelder and Piaget, 1958), 54 percent failed to separate and control variables on the bending rod task (Inhelder and Piaget, 1958), and 47 percent did not exhibit conservation of displaced volume (Karplus and Lavatelli, 1970). These tasks were all administered by personal interviews. In a Canadian study Wheeler and Kass (1977) reported, after group administration of a battery of four tasks involving written answers, that only 43.5 percent of a sample of 168 high school subjects were judged to be formal thinkers. These results are typical of the literature in general. Whether they reflect inherent intellectual structural limitations, as Piaget's theory suggests, is open to question. An extensive study by Karplus and others (1975) found differences between national groups of subjects in a sample of over three thousand subjects drawn from the United States, Great Britain and five European countries on tasks testing proportional reasoning and controlling variables. Differences between countries were not consistent over both tasks. In view of the wide variations the characteristics of the samples, the authors' conclusion that teaching can have some influence on the development of reasoning is not necessarily justified. However, the particular differences in educational and instructional practices in these different countries may certainly bear investigation for any light they may shed upon the development of reasoning. Certainly the results cast some doubt upon the notion of structural limitations, and indicate the reasonableness of recent studies designed to elicit the effect of specific training on the development of formal operational thinking.

Developmental Level and Achievement in Science

The previous section addressed the low incidence of formal thought exhibited by the high school population. In the present section the relationship between this deficiency and the achievement of this population in their science courses is addressed.

Several authors have analyzed typical school science curriculum materials in terms of why they appear to demand formal operational thought on the part of the learner (Ingle and Shayer, 1970; Shayer, 1973; Herron, 1975; Karplus, 1977).

Ingle and Shayer have presented an analysis of Nuffield 'O' level chemistry in terms of Piaget's stages and also I.Q. If their analysis is correct, the chance of success in this course for the average ability child is remote. Of interest to the present study is the conclusion that the mole concept, which these students meet at age thirteen or fourteen, is a source of particular cumulative difficulty because it is seen as the main integrating concept of the course during the third of the five years of the course. In a later paper Shayer (1973) again emphasizes the formal operational nature of chemistry and suggests its conceptual demands to be considerably higher than those of physics or biology at similar school levels. He strongly suggests that content requiring more than concrete operational thinking will not be meaningful to the child of average intelligence until he is sixteen years of age. Herron (1975) compiled a list of 16 commonly expected competencies which concrete operational students could be expected to exhibit and contrasted each with the formal operational extension normally required by the science curriculum materials. For example, he suggested that the concrete operational thinker can conceive of atomic weight as the mass of the Avagadro number of atoms. However, according to Herron the same subject cannot conceive of atomic weight as the ratio of the mass of one atom to the mass of some other atom which is selected as standard. Herron follows his analysis with a suggestion that we focus on a concrete approach to chemistry where possible, although he is simultaneously concerned that we should not neglect to attempt to develop the student towards formal operational thought. This view was also expressed by Howe (1974), who suggested that "Teachers should not wait for students to become formal operational. It may never happen." Howe also concurs on the need for a concrete operational mode for most instruction in school science. An interesting comment on this suggestion is made by Munby (1978) who notes that the use of concrete referents in science instruction may be quite justified and is certainly found in Piaget's writings, but cannot be derived from Piaget's theoretical framework. Karplus (1977) notes that some commonly taught concepts, such as chemical bond and gene, demand formal operational thought. Other concepts, such as acid or cell, may be presented in either a formal or a concrete operational mode.

The apparent lack of congruence between the intellectual developmental characteristics of high school science students and the curriculum content to which they are exposed leads to the question of whether an empirical relationship exists between the developmental level of the learner and his achievement in science. As yet such evidence is scarce.

One such study by Sayre and Ball (1975) identified a relationship between developmental level and science achievement as measured by grades for some sub-samples of a randomly drawn sample of junior high students and a second randomly drawn sample of senior high students. A total of 419 subjects were drawn from grades seven to twelve. Developmental levels were determined after administering five tasks in an interview format. These were Karplus and Peterson's Mr. Tall and Mr. Short task, Inhelder and Piaget's pendulum, equilibrium in the balance and chemical combinations tasks, and a task requiring valid argument with respect to a syllogism. Subjects who succeeded on four or five tasks were judged formal operational. Those who failed at least two tasks were judged non-formal operational. For both samples formal operational students received significantly higher grades (at the 0.01 level) than non-formal operational students. The actual point-biserial correlations for these groups, however, were only moderately high, being 0.33 and 0.45 respectively. Moreover, for the grade seven general science and the grade twelve physics sub-samples, no relationship between developmental level and achievement was found. However, Herron (1976b) notes that the extreme proportion judged formal in these two groups, 8.6 percent and 80.7 percent respectively, mitigates against finding significant correlations. Sayre and Ball concluded that secondary science instruction should be structured around the cognitive developmental level of the students involved. Herron (1976b) notes two conflicting interpretations of this. Either there is a need to restructure the science curriculum to eliminate the need for formal thought, or alternatively instances of formal thought should grow out of and be based on concrete experience. He suggests the latter to be more appropriate.

A study by Lawson and Renner (1975) examined the relationship between learner developmental level and acquisition of concepts classified as concrete and formal. The sample consisted of 51 grade ten biology students, 50 grade eleven chemistry students and 33 grade twelve physics students. Each of these groups represented intact classes. The developmental level of each subject was assessed by application of six tasks which were administered together in individual interviews with each subject. Responses were scored on a scale from zero to five with the maximum for each task varying according to the developmental level implied by the task. Each subject was placed on a developmental scale ranging from early concrete to late formal after consideration of his total score on the six tasks. After the developmental data were collected a multiple choice test containing only questions requiring comprehension and application was administered in the relevant subject area. It was considered that understanding of the concepts involved would best be tested by questions at the two levels indicated. The concepts tested were classified as concrete-operational if their meaning could be developed from first-hand experience with objects or events, and formal-operational if their meaning required the use of a "postulatory-deductive system." Two separate one-way analyses of variance were conducted to determine the relationship between developmental level and proportion of correct responses to the concrete-operational and the formal-operational questions, respectively. Significant differences were found.

The authors concluded that:

> . . . concrete-operational subjects are unable to develop understanding of formal concepts. Also, support is demonstrated for the other major premises of the study: concrete-operational subjects are able to demonstrate understanding of concrete concepts, and formal-operational subjects are able to demonstrate understanding of both concrete and formal concepts.

Two aspects of these findings merit comment. Firstly, the results do not show that concrete-operational subjects are unable to develop understanding of formal concepts. What can legitimately be claimed is that in this study they did not. Secondly, none of the group, including those classified 'formal-IIIB,' was successful on more than half the formal questions. This outcome is of some importance to the present study. It is possible to speculate several reasons for this. One of these may be derived from the learning hierarchy model. In particular, failure by both formal and concrete individuals might be related to the absence of specific content-related prerequisite capabilities.

A slightly different approach to the question of the relationship between developmental level and achievement in science was taken by

Lawson and Nordland (1977). The basic hypothesis of this study was that level of conservation reasoning as a necessary prerequisite for rational thought should be related to achievement in science. Conservers should exhibit greater acquisition of formal-operational concepts. The sample was composed of 20 male and three female subjects enrolled in an elective high school biology course which used the BSCS Blue Version as a textbook. Most of the subjects were considered to be of above average intellectual ability.

The procedure used was to administer the conservation tasks at the beginning of the course and then to administer six tests derived by selecting 20 to 30 items from BSCS examination materials over the course of the ensuing semester. When selecting the test items care was taken to categorize them first as requiring concrete or formal thought and to choose approximately equal numbers of items of each category. At the end of the semester total scores on concrete and formal items were obtained for each subject. The significance of any differences between the four developmental groups from early concrete to late formal on each of these totals was determined by one-way analysis of variance both before and after correcting for guessing. No significant differences were found at the 0.05 level but differences were found at the 0.10 level. The rationale for choosing this level of significance is not clear. Although it could be argued that the reason is to increase the power of the test, this result could have been more appropriately achieved by using a larger sample. Despite this, examination of the form of the data certainly suggests that responses to certain conservation tasks were related to achievement of formal-operational concepts. However, even those subjects who were successful on all three tasks could average only 22.4 percent success on the formal operational biology questions, after correcting for guessing.

The Lawson and Nordland study just discussed focused entirely upon responses to conservation tasks for the developmental level testing involved, yet no attempt was made to relate the responses to these tasks directly with responses to test items requiring conservation. Shayer and Wharry (1974) suggest that it may be most reasonable to focus upon developmental level tasks which relate to particular schema directly related to concepts to be learned. An example of this approach is a study by Wheeler and Kass (1978) who reported that the ability to combine variables as evidenced by combined performance on Piaget's combination of colorless solutions task and a similar task devised by the author was a highly significant predictor of achievement on a test of chemical equilibrium, a concept that is clearly related to the particular schema tested by the developmental tasks. It seems possible that such a result is more related to the specific prerequisites suggested by the Gagnéan model than to the more generalized intellectual competencies represented in the Piagetian model.

Another study by Wheeler and Kass (1977) attempted to relate subjects' proportional reasoning ability to achievement in introductory chemistry. The sample consisted of 168 grade ten students enrolled in an introductory chemistry course. At the beginning of the course a group of four neo-Piagetian tasks involving written answers was administered, followed by a test of general proportional reasoning.

During the chemistry course a test of proportionality in chemistry was administered in four parts in regular classroom periods when the relevant chemistry had been covered. These sub-tests related successively to chemical nomenclature and the writing of formulae, chemical reactions, the mole concept, and gravimetric stoichiometry. The items on these sub-tests were constructed to be analogous to the items in the general proportionality test. At the end of the course a general chemistry achievement test was administered, and also a second general proportionality test which was equivalent to the first. Not surprisingly the best correlations with chemistry achievement were the chemistry proportionality sub-tests. The neo-Piagetian reasoning test and the general proportionality test, although showing moderate correlations (0.48, 0.46, 0.41) with achievement, did not add significantly to a regression equation with chemistry achievement as criterion once the scores from the chemistry proportionality sub-tests had been entered. These four sub-tests accounted for 63.6 percent of the variance. However, developmental level, categorized by responses to the four neo-Piagetian tasks was significantly (0.001) related to chemistry achievement.

Determination of Developmental Level by Group Testing

For the investigator interested in the relationship between curriculum and developmental level the use of large numbers of subjects is desirable. However, use of Piaget's individual interview technique raises the problem that samples are likely to be small because of the time needed for interview. An alternative is to use group-testing methods, with written responses. Lawson's (1978) contention that what is fundamentally important is what is measured rather than how it is measured is a reasonable argument in favor of such alternative procedures, although it must be recognized that the flexibility of the interview technique is lost when group testing is used. Verbal limitations on the part of the subject and the inability of the investigator to probe the subject's responses are in turn limitations upon this method. The change of format may also change the nature of the subject's response. For example, Bruner (1966) showed that a format allowing subjects to be perceptually distracted in a task testing conservation of liquid suggested a lower proportion of conservers.

A similar result was obtained by Karplus, Karplus and Wollman (1974) on a ratio task. The task involved calculating the height of 'Mr. Tall,' being given the height of 'Mr. Short,' and the ratio of their heights. The difference between the two administrations was that for the first group of children both Mr. Tall and Mr. Short were present, whereas in the second administration Mr. Tall was absent. Although the second administration might be held to involve a more abstract task the elimination of the perceptual stimulus of Mr. Tall's presence appeared to produce much fewer perceptually-bound responses. The result was a substantial shift towards a higher level of operational thinking. Despite these potential limitations progress has been made towards effective group-testing procedures, as is apparent from the reports of several such studies which are now described.

Raven (1973) produced 'Raven's Test of Logical Operations' which was divided into three parts. The first part tested classification and seriation, the second part tested logical multiplication, and the third part tested probability and correlation. Each item consisted of a problem presented pictorially and a brief written question. In each case the subject was required to select the correct answer out of three alternatives. The test was validated by expert judgment and empirical testing, with a panel of four Piagetian experts exhibiting 100 percent agreement that the items measured the logical operations for which they were designed.

Shayer and Wharry (1974) described the use of a number of tasks, some related directly to Piaget's work. These authors considered it to be particularly important that the tasks be seen as tasks not tests; that they should be relevant to the curriculum structure of the school; that pupils of a wide ability range should be able to produce answers which they, at least, felt to be satisfactory; and that they should be designed primarily to measure the mental processes of the child in relation to his immediate area of study. It appears, therefore, that Shayer and Wharry are concerned about the availability of appropriate schema at the appropriate time as well as being concerned about diagnosis of developmental level in general. In their study questions involved both verbal and written instructions with demonstrations where necessary. Answers were written.

A study by Rowell and Hoffmann (1975) focused directly on the use of two Piagetian tasks, 'chemical combinations' and 'the pendulum task,' deliberately chosen to represent different content but the same underlying intellectual structure. Subjects performed the tasks individually from written instructions and were asked to record their responses in writing. The need to record all manipulations, reasons, and conclusions as fully as possible was stressed. Scoring of the responses followed Piaget's criteria closely. Inter-judge reliability was extremely high (0.96 and 0.92 for the two tasks). A significant correlation between responses to the two tasks was obtained. Generally, where a subject's response was different in the two tasks his developmental level appeared higher on the pendulum task, suggesting that it measures an earlier level of formal operational thought than does the chemical combinations task. The authors concluded that a group administered test of developmental level in which the subjects manipulate materials and in which written responses are scored by independent raters such as teachers is feasible. The general nature of the results was consistent with those obtained from studies using interview techniques. However, the conclusion cannot be drawn necessarily that results from the two techniques are equivalent.

A comprehensive attempt to develop and validate a classroom test of formal reasoning has been reported recently by Lawson (1978). Lawson suggests that the several attempts which have been made to produce strictly paper and pencil measures of formal reasoning ability depart too far from the clinical setting demanded by Piaget, a point already noted by Shayer and Wharry. Essentially, he argues that a set of written questions represents a test rather than a set of tasks. Moreover, as has been indicated previously, extent of exposure to the concrete materials of the task may change the nature of subjects' responses. Hence all of the tasks in Lawson's test involved a demonstration using some physical materials or apparatus. Subjects responded by selecting one of a number of alternative answers and then explaining the reasons for their choice. The test included fifteen tasks, most taken from other well documented studies. Groups of tasks tested for isolation and control of variables, combinatorial reasoning, probabilistic reasoning, and proportional reasoning. A number of the tasks were taken directly from Inhelder and Piaget (1958). One task testing conservation of weight was included.

The resulting test was administered to a total of 513 students enrolled in grades eight to ten. The sample was randomly drawn from a middle socioeconomic population. Testing took 75 to 100 minutes, and was administered over two testing periods to intact classes. Importantly, in order to determine construct validity of the test in relation to Piaget's own procedures, a sub-sample of 72 subjects were randomly selected and individually administered a battery of four of Piaget's tasks by interview. Such a procedure is perhaps more defensible than that of Raven, who relied upon expert judgment alone to determine construct validity of his test of logical operations. Surprisingly, scores on only two of the four tasks were used to obtain a correlation between scores on the group administered test and on the interview administered test. One of the excluded tasks, conservation of weight, is a measure of concrete operational thought and the other, conservation of displaced volume, is considered an indicator of early formal operational thought. The rationale was advanced that as neither represented fully formal operational thought, they should be excluded. Such reasoning appears illogical in that both tasks were used on the written test. A second inconsistency is that the interview tasks were scored on a four-point scale according to the perceived developmental category of the response, while the responses to the tasks in the group administered test were scored right if the correct choice was made together with an acceptable reason and otherwise scored wrong. The question must be asked, despite very high correlations and encouraging factor analysis results, whether the same pattern would have emerged if the tasks in the group test had been scored according to the developmental level of the response and if a free response rather than a forced response format had been used. As Lawson (1978) notes the study by Rowell and Hoffman indicates the feasibility of such an approach. In summary, it is suggested that Lawson's study offers further support to the usefulness of group testing of developmental level, but is perhaps less definitive than he suggests.

Renner (1977) has also attempted to produce a group administered test involving written responses. He, too, attempted to validate the test by relating subjects' responses on it to their responses on standard Piagetian interviews, a procedure which seems desirable in future similar studies. Despite the similar intent, the nature of Renner's test was very different to that of Lawson. Essentially, it takes the form of groups of written incidents to which the learner must respond. The content of each incident was carefully chosen to represent science, but no particular factual knowledge was required. The reading level of the incidents was deliberately well below the expected reading ability of the sample with the maximum reading level required by any incident being less than seventh grade. The subjects for both the pilot study and the final study were from grades ten, eleven and twelve science classes. The incidents, and the four Piagetian tasks used to aid their validation, focused upon combinatorial logic, proportional reasoning, and separation and control of variables. In the final study, two sets of three incidents were used, reflecting these three schema. Scores for each of the incidents and also for the Piagetian tasks were derived by consideration of the reasoning patterns exhibited. To provide a basis for relating an individual's score on a set of incidents to an equivalent score on the set of Piagetian tasks a stepwise regression equation was derived. The total score on the Piagetian tasks was used as criterion and the scores on separate incidents were successively added as predictors. The best predictor accounted for 54 percent of the variance. Addition of two more incidents increased this to 62 percent, after which no further increase was observed. The incidents representing the best three predictors represented use of proportional reasoning, separation and control of variables and combinatorial logic, respectively.

Renner contends that removing the element of social transmission by substituting written incidents for interview-administered tasks is undesirable, and probably contributes to a lowering of the correlation between the tests. However he claims, justifiably in principle at least, that a test which can be quickly and reliably administered and scored is usable and likely to be used by the classroom teacher as a useful introductory guide to the developmental level profile of his class. He estimates that the test takes twenty-five minutes to administer to a group and about five minutes to score for each individual. Conversely, he claims that most teachers would experience difficulty administering and scoring a group of tasks by an interview technique and that the time required would be prohibitive.

Renner's study is exceptionally well done. It seems possible that a higher correlation between Renner's incident battery and the Piagetian battery might have been obtained if the tasks had involved some element of demonstration. However, the study adds substantially to the body of evidence that it is possible, with reasonable justification, to replace Piagetian interview techniques with written tests.

A number of other attempts to apply large scale testing, which almost of necessity implies written answers, have been made. Collectively, they, too, add to the weight of evidence that such procedures are feasible. Although often no formal attempt has been made to validate the tests against results obtained from interview data, they are supportive in that typically the proportions of subjects placed in each Piagetian stage are typical of those found when similar subjects elsewhere have been interviewed (Karplus, 1977; Hobbs, 1975; Wheeler, 1976).

Summary

The present chapter has described a number of studies which indicate that a substantial proportion of high school students are not formal operational, that developmental level correlates significantly with achievement in science, and finally that group testing procedures offer a useful way of identifying the developmental levels of individuals.

Chapter 5

DESIGN, INSTRUMENTATION AND PROCEDURES

In the identification of learning hierarchies the steps involved generally include generation of an hypothesized hierarchy, development of suitable test questions for each element of the hierarchy, design and implementation of appropriate instructional and testing procedures, use of these procedures with a selected sample, and analysis of the results. To these the present study adds an investigation of the relationship between performance on selected developmental tasks and achievement of the elements of the hierarchy. Hence, selection and administration of appropriate developmental tasks was also involved. This chapter describes the practical aspects of each of the foregoing steps in the present study, as well as discussing the rationale for each decision made.

Construction of the Hierarchy

In chapter one it was indicated that a fundamental reason for requiring students in introductory chemistry courses to learn the mole concept is to enable them to consider masses of chemical substances in terms of the relative numbers of particles represented. In the present study the ability to do this was taken as the key skill in the development of the mole concept. Hence this was chosen as the superordinate skill for the hierarchy hypothesized for the study. The hypothesized hierarchy was derived by successively asking the question 'what should the learner be able to do first if he is to learn this skill?' Each skill so derived was subjected to the same question, and this process continued until a skill was reached which appeared to represent the beginning of the concept. It was recognized that the superordinate skill in this hierarchy is in turn subordinate to other skills in the many applications of the mole concept, as for example in stoichiometric calculations, and that a much larger hierarchy could be hypothesized. However, restriction to the hierarchy actually hypothesized was made for several reasons. Firstly, there are practical limitations to the amount of time available in schools for research purposes. Within the design of the present study even restriction to the small amount of content involved requires three quizzes and a final test in addition to the 'Test of Developmental level' and a second test related to developmental level, the Skemp Test (Skemp, 1960). Secondly, it is suggested that more may be gained by a closer analysis of a small area of content than by a more gross analysis of a larger area. In this context it is argued that interpretation of the tests of relationship between the elements of the hypothesized hierarchy will be improved by using more questions per skill than by expanding the content and using less questions, given limited time for testing. This is consistent with the comment made in Chapter two that the present trend appears to be toward investigation of small hierarchies. Thirdly, the skill chosen as the superordinate skill represents the fundamental meaning of the concept. It was considered that if an hierarchy leading to the learning of this skill could be identified, further studies might then be appropriate leading to potential hierarchies in areas of application of the mole concept.

Application of the task analysis procedure described followed by modification after the first pilot study led to the identification of the eight skills. A description and illustrative example of each skill follows:

A. Calculate the masses of different elements or compounds containing the same or proportionate numbers of atoms or molecules.

For example, 'Calculate the mass of ammonia (NH_3) containing twice as many molecules as 32 grams of oxygen (O_2).'

B. Convert the mass of an element or compound to the number of atoms or molecules present, and vice-versa.

For example, 'How many molecules are there in 23 grams of nitrogen dioxide (NO_2)?'

C. Determine the relative number of atoms or molecules present in given quantities of elements or compounds.

For example, 'Consider 10^{12} atoms of calcium. How many atoms of carbon are needed to give the same number of moles as this number of atoms of calcium?'

D. Convert a given number of moles of an element or compound to the number of atoms or molecules present, and vice-versa.

For example, 'How many moles of the named compound are there in 6.02×10^{23} molecules of carbon monoxide (CO)?'

E. Calculate the mass of an element or compound containing the same number of moles as a given mass of another element or compound.

For example, 'How many moles of magnesium would weigh the same as six moles of carbon?'

F. Convert a given mass of an element or compound to the number of moles represented, and vice-versa.

For example, 'How many moles of the named compound are represented by 88 grams of carbon dioxide (CO_2)?'

G. Apply the definition of mole as it relates to the Avogadro number of atoms or molecules and the molar mass of an element or compound.

For example, 'Which of the following represents one mole of hydrogen chloride (HCl)?

(a) $\frac{36.5}{12}$ grams   (b) 36.5 grams

(c) $\frac{6.02 \times 10^{23}}{36.5}$ grams   (d) $36.5 \times 6.02 \times 10^{23}$ grams'

H. Identify and apply the definition of molar mass as a ratio.

For example, 'If the molecular weight of a gas is 33, what is the molecular weight of a gas whose molecules are three times as heavy?

(a) 3   (b) 11

(c) 99   (d) some other number'

The arrangement of skills A to H in their hypothesized hierarchical relationship is shown in Figure 6.

The reasonableness of this hierarchy was discussed with five science educators and the ten chemistry teachers who were involved either in the study or one of the pilot studies. In general the hierarchy was given approval by each group but a number of the teachers

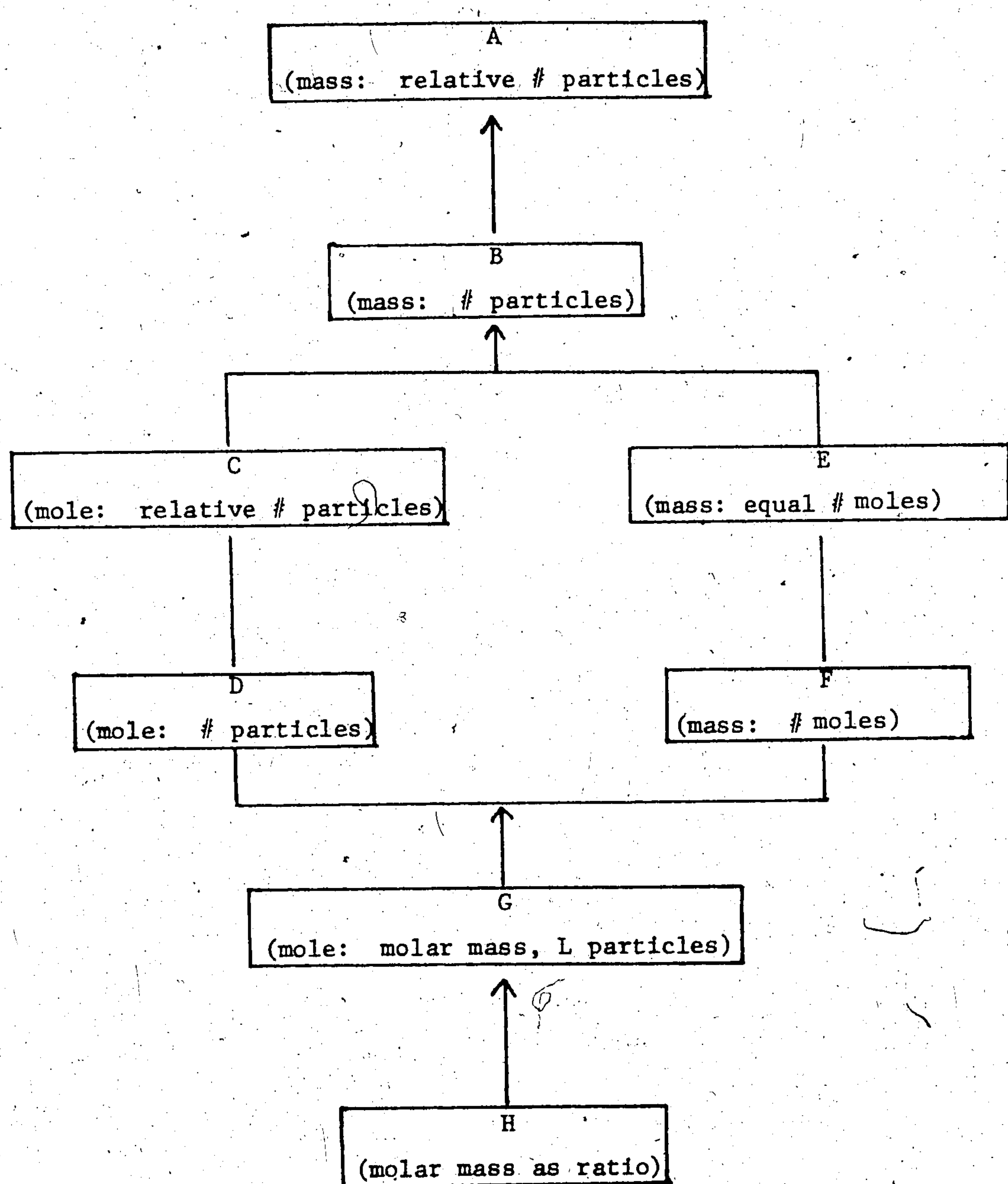


Fig. 6. The hypothesized hierarchy

suggested that different responses might be obtained when questions involving chemical compounds are used than when questions involving chemical elements are used. This suggestion was incorporated into the study by the investigator by including parallel test items for elements and compounds in the tests for each skill in the hypothesized hierarchy, allowing testing of the following two possibilities:

1. Does the hypothesized hierarchy or a modification of it represent a valid learning hierarchy?

2. Within such an hierarchy does some subdivision of generalized skills exist which allows for hierarchically arranged sub-skills? A parsimonious way of expressing this possibility is shown in Figure 7.

Several additional comments relating to the hierarchies represented in Figures 6 and 7 are appropriate at this juncture. Firstly, the sequence of development of the intellectual skills involved is consistent with Gagné (1970a) in that Skills G and H represent the development of discriminations and concepts while Skills C, D, E and F represent the direct application of rules relating to these concepts and Skills A and B represent the use of combinations of these rules, which is the use of higher-order rules.

Secondly, an important question which must be answered in constructing any hierarchy relates to the step size between adjacent skills. If the step size is too large important information may be missed. On the other hand, if it is too small the need to set more test questions either decreases the breadth of content which can be covered or demands more testing time. A particular problem in this respect is the effect of numerical competence, which appears to be a

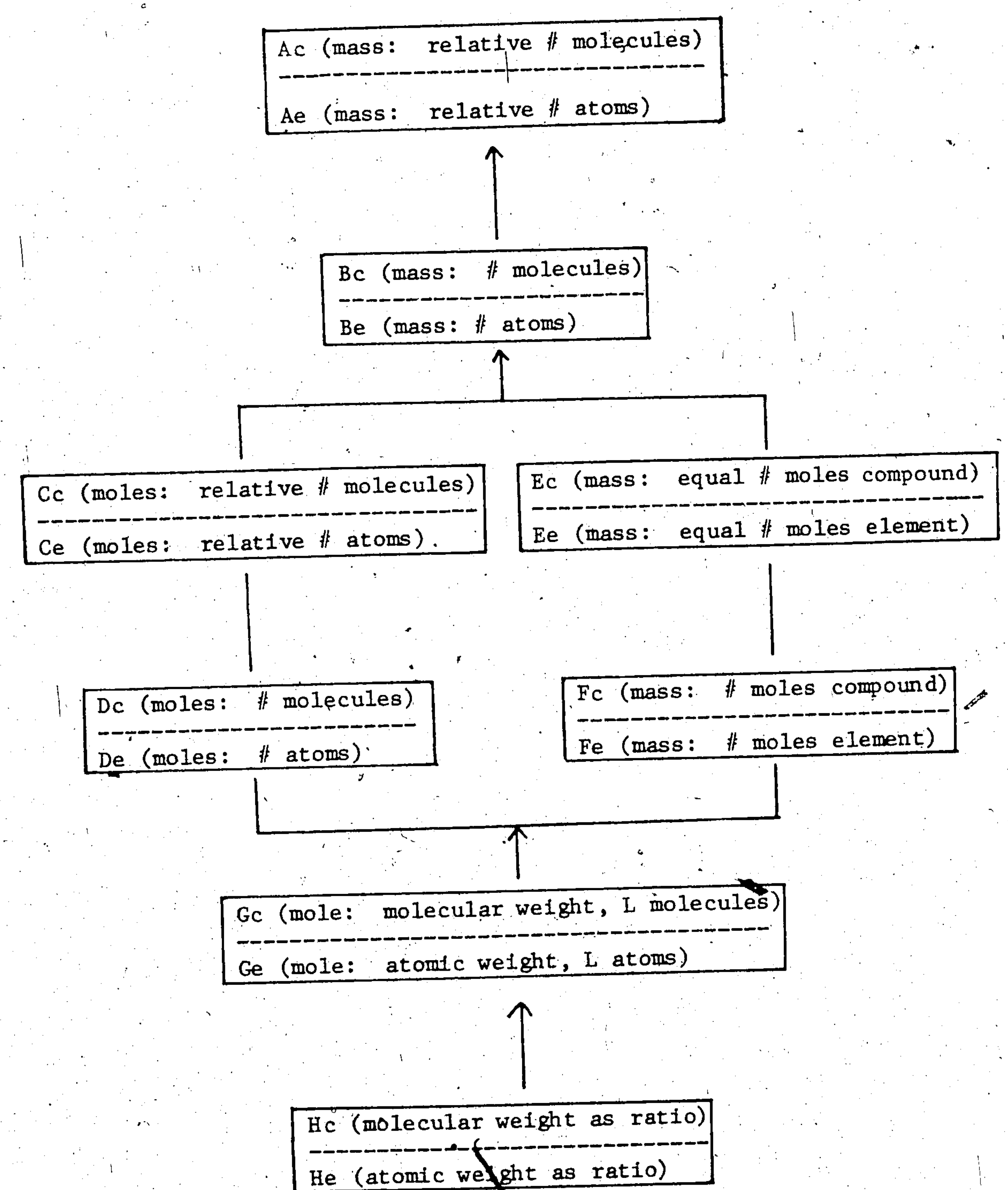


Fig. 7. Modified hypothesized hierarchy

major factor in the solution of chemistry problems (Denny, 1970; Dence, 1970; Good and Morin, 1978). Okey and Gagné's (1970) learning hierarchy for solubility product calculations includes a large proportion of skills representing numerical competence. However, it is argued here that too substantial an emphasis on numerical skills may distract from the identification of conceptual relationships in the understanding of the chemistry involved. Hence, in the present study skills requiring only numerical competence were avoided and the need for numerical competence in test items was minimized. The use of mainly free-response test items hopefully enables identification of numerical difficulties as opposed to difficulties with the chemical concepts themselves.

Thirdly, two troublesome questions encountered in the construction of the hypothesized hierarchy may be mentioned. The first relates to whether to include the calculation of actual numbers of particles in given mass or mole quantities in the hierarchy, given that the superordinate skill requires the comparison of different masses of substances in terms of the *relative* numbers of particles present. The investigator rationalized the inclusion of skills involving actual numbers of particles as providing a more concrete route to the superordinate skill. Several of the participating teachers argued that calculations involving actual numbers of particles were not necessary. The authors of the two most recent revisions of Chem Study (O'Connor et al., 1968; Parry et al., 1969) argue that this skill is necessary. Parry and his co-authors note: "We have stressed the significance of the number of molecules in a mole and have extended this to any particle. . . . This use was found to be of great help in the original Chem Study." O'Connor and his co-authors suggest: "Make sure the students know the relationship between molar mass and Avogadro's number." The same point was made by Herron *et al.* (1977) on the basis of a theoretical analysis of the mole concept. Moreover, data from the first pilot for the present study was suggestive of a separation of Skill D representing inter-conversion of moles and *actual* numbers of particles from Skill C representing identification of *relative* numbers of particles in different mole quantities. The calculation of actual numbers of particles in mole quantities was therefore retained in the hypothesized hierarchy.

A second problem relates to the two lower skills in the hypothesized hierarchy. Gagné (1969b) has noted that questions asking for verbalization of knowledge (definitions) tend to produce anomalous connections in hypothesized hierarchies. For example, use of Kolb's (1967) hierarchy leading to the ability to apply certain math skills was not successful in promoting acquisition of the superordinate skill, suggesting that the hierarchy was not valid. According to Gagné (1969b) and White (1973) the cause of this invalidity was the inclusion of many 'skills' requiring only verbalization of a definition. A learner who can repeat a statement does not necessarily comprehend the meaning underlying the statement and may not be able to apply it. Conversely a learner who can apply what is represented by the statement may not be able to adequately verbalize the statement. Hence, for these two reasons the inclusion of verbalized statements in hypothesized learning hierarchies appears to be likely to produce spurious connections. Consequently, the hierarchy hypothesized in the present study attempts to test understanding of relevant definitions by requiring identification and application of them rather than verbalization.

White (1974a) has suggested the following model for identification and validation of learning hierarchies:

1. Define, in behavioral terms, the element which is to be the pinnacle of the hierarchy.

2. Derive the hierarchy by asking Gagné's question ("What must the learner be able to do in order to learn this new element, given only instructions?") of each element in turn, from the pinnacle element downward. Include all connections that seem reasonably possible, since the validation process can only destroy postulated connections, not create them. Avoid verbalized elements, they can be included in the instructions.

3. Check the reasonableness of the postulated hierarchy with experienced teachers and subject-matter experts.

4. Invent possible divisions of the elements of the hierarchy, so that very precise definitions are obtained.

5. Carry out an investigation of whether the invented divisions do in fact represent different skills. One way of doing this is to write two or more questions for each division and give them to a sample of Ss. Wherever any Ss are observed to answer correctly the set of questions for one division, while answering incorrectly the set for another, the divisions are taken to be separate skills. White has given a description of the practical arrangements of such an investigation.

6. Write a learning program for the elements, embedding in it test questions for the elements. The questions for an element should follow immediately after the frames that teach the element. There must be two or more questions for each element to allow for an estimate of their reliability.

7. Have at least 150 Ss, suitably chosen, work through the program, answering the questions as they come to them.

8. Analyze the results to see whether any of the postu-

lated connections between elements should be rejected. A suitable test of a hierarchical relationship has been developed by White and Clark. The hypotheses compared in the test are Ho: the proportion of the population from which the sample was drawn who can learn higher element without the lower element is zero; and Ha: the above proportion is greater than zero. The test provides estimates of the probabilities of the observed results given that Ho is true or given specific values of the proportion under Ha.

9. Remove from the hierarchy all connections for which the probability under Ho is small, say 0.05 or less.

Steps one to five of White's model were applied with little change in the present study. As has been indicated earlier, two other current alternatives to the White and Clark test will be applied to the analysis of results. Hence the design of the present study involves modification of steps eight and nine of White's model. Further, with regard to steps six and seven which represent the instructional and testing components of White's model substantial differences exist in the present study. These will now be discussed.

White's model is typical of most learning hierarchy research in the form of its instructional component. The recommended format for instruction is the use of programmed materials. Hence the potentially confounding influence of the teacher is eliminated. Clearly this is a source of satisfaction with respect to considerations of internal validity. However, although learner progress through an hierarchy may be dependent upon instruction, Gagné (1973, pp. 21, 22) may be interpreted as suggesting that the *structure* of any particular learning hierarchy is independent of instruction. This independence between the structure of the hierarchy and instructional practice is important to the conduct of the present study. To illustrate this point further consider a particular group of learners. Given appropriate instruction all may progress completely through the content represented in a particular hierarchy. Alternatively, if instruction is inadequate they may progress only partly through the hierarchy. What should not happen if an hierarchy really exists is that many learners exhibit any later skill(s) having failed to exhibit any particular earlier skill(s) regardless of instructional quality or practice. Learning hierarchies should therefore be amenable to investigation, and capable of generalization, outside the bounds of the programmed instruction format. Moreover, it may be argued that the sequence of a particular programme might induce the existence of a hierarchy that is consistent with that particular instructional sequence, but which lacks validity in general terms. A variety of approaches such as one would expect to find from a number of teachers not constrained by any particular set of instructional procedures might produce an hierarchy with acceptable internal validity and greater external validity. In the present study no attempt was made to interfere with the preferred instructional practices of the teachers involved.

A second important departure from White's recommended procedures relates to an attempt to add evidence of positive transfer to the psychometric evidence suggested by White. The need to provide evidence of transfer was discussed in Chapter two. The particular method employed was chosen according to the rationale that a group of subjects who failed to show evidence of learning subordinate skills at the time of instruction should, given appropriate remediation before instruction progressed to the next skill, show greater achievement of related superordinate objectives than a similar group not receiving remediation.

A third departure from White's recommendations is that testing of subjects' mastery of skills was carried out after instruction was complete as well as during the instructional period. Indeed, for the reasons which follow it is argued here that testing after instruction is more appropriate than testing during instruction, and the major reason for testing during instruction in the present study is to provide a particular means of estimating the existence of positive transfer of learning from the lower skills to the upper skills in the hypothesized hierarchy. White (1973) suggests that testing after the instructional process is complete is likely to be misleading in that anomalies may be produced because subordinate skills may have been learned but forgotten by the time of a final test. However, it is also possible that learning of 'missed' skills may take place as a result of initial testing or for some other reason prior to the testing of later skills. Hence, it is at least equally conceivable that White's method of testing during instruction may produce anomalies. Moreover, the possibility of short-term memory effects also suggests that immediate testing may produce some anomalies. The design of the present study in which parallel tests were administered during and after instruction allows an empirical comparison of the form of the hierarchies implied by data collected at these two times. For this purpose the sub-group of students not receiving remediation forms an appropriate sample.

In summary, the procedures used to derive the hypothesized hierarchy for the present study are similar to those suggested by steps one to five of White's model. However, the procedures used to determine the validity of the hypothesized hierarchy are substantially different to the remaining steps of White's model. In particular, teacher instruction rather than programmed instruction is involved, testing is carried out after as well as during the instructional period, remediation of missed skills is used to provide a test of positive transfer and several statistical methods are applied to determine the validity of the hierarchy. The actual design involved is illustrated in Figure 8. In addition to being concerned with the identification of a learning hierarchy relating to the mole concept the present study is also concerned with identifying the relationship between learner developmental level and acquisition of the skills which compose the hierarchy. The placement of the developmental level testing is therefore also included in Figure 8.

Sample

The sample consisted of 269 grade ten students enrolled in introductory chemistry programmes in four senior high schools under the jurisdiction of the Calgary Board of Education. The mean age was 16.3 years, with a standard deviation of 0.3. There were 133 boys and 136 girls. Eleven classes and eight teachers were involved. The intake of the schools represented a variety of socioeconomic backgrounds. In general, the sample appears quite representative of North American urban areas.

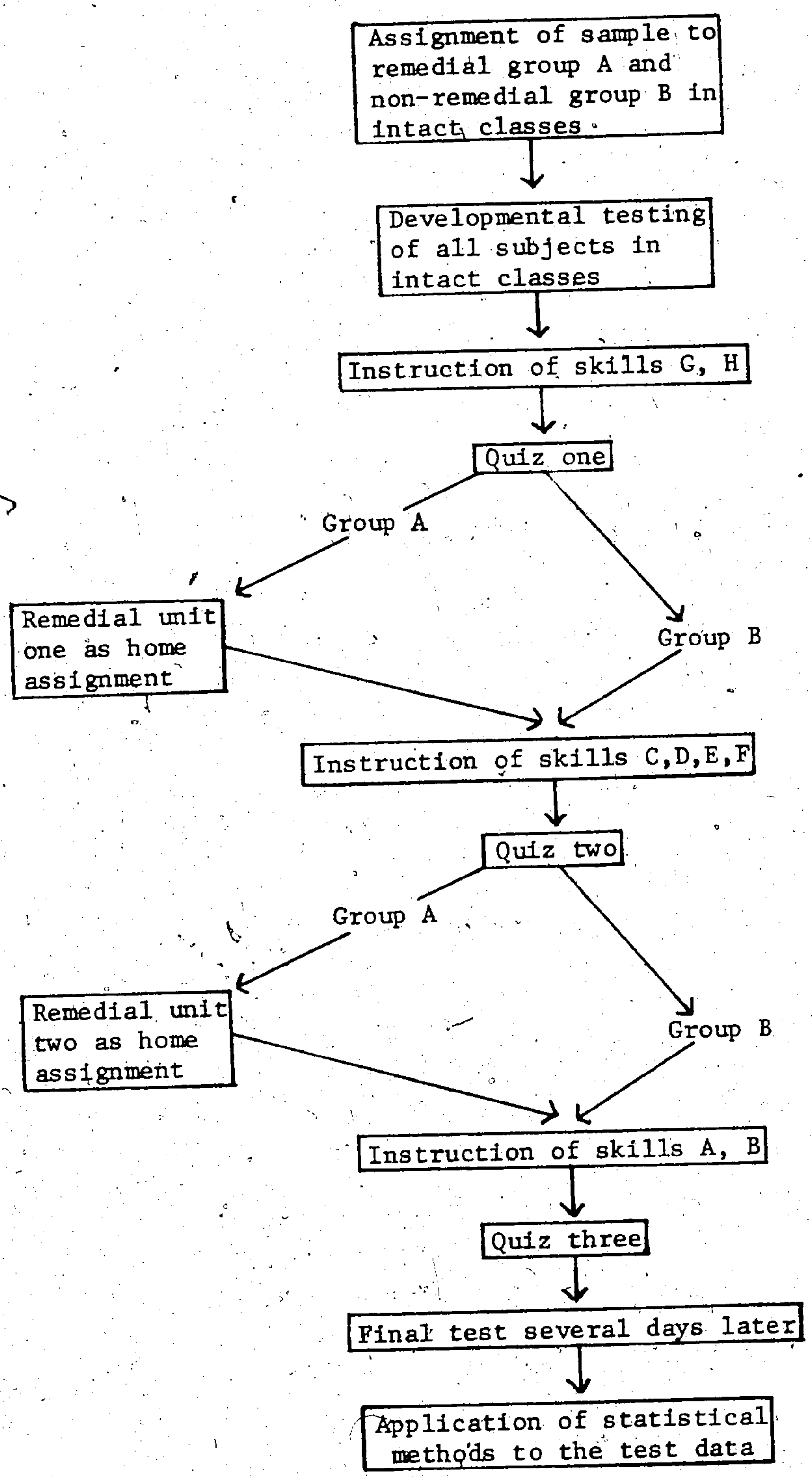


Fig. 8. Design of experimental procedures

Procedures

The procedural aspects of the study include procedures relating to instruction, testing and remediation, respectively.

It has already been suggested that the structure of a learning hierarchy should be independent of instruction. Hence, the teachers involved in the present study were encouraged not to depart from whatever constitutes their usual instructional practices. The only exceptions to this were the requirements for each teacher to administer the chemistry tests provided by the investigator and those teachers involved with classes designated as remedial to administer the remedial units as directed by the investigator. Each teacher was required to follow a course of study based upon the use of the original Chem Study text (Pimentel, 1963). However, discussions with the teachers yielded the information that the mode of instruction varied between schools and to a lesser extent between teachers in the same school. In three of the schools instructional practice could best be characterized as conventional, involving basically teacher exposition accompanied by student laboratory activities; in the fourth school the two teachers involved applied a team-teaching approach in an individualized setting. In one school one of the two teachers involved applied the use of algorithmic procedures to mole calculations while the other did not favor or use such procedures.

A few days prior to the commencement of instruction relating to the content of the hierarchy the Test of Developmental Level and the Skemp test were administered to intact classes by the investigator. In three of the schools these tests were administered in one double-period sitting while in the fourth school it was necessary to use two single periods. In each case the Test of Developmental Level was administered first, and sufficient time was available for all subjects to complete each test.

In order to determine the existence and strength of transfer of learning from the subordinate skills in the hierarchy to related superordinate skills the sample was divided into two sub-groups before the commencement of any testing or instruction procedures. One of these sub-groups was designated as remedial and each subject in this sub-group was required to complete two remedial units as take-home assignments between testing of the skills involved and instruction of the next skill. The intent of this procedure was to determine if successful remediation of inadequately learned subordinate skills was related to acquisition of related superordinate skills. Hence, before the commencement of testing or instruction subjects were assigned in intact classes to the remedial and non-remedial groups. This was done by random selection of teachers who were to administer the remedial treatment. Where a teacher taught more than one class each class was treated similarly. Each subject in the remedial sub-group was required to complete and return both remedial units whether or not the subject had shown mastery of all skills in the preceding quiz, although for the purpose of analysis subjects were considered remedial only for those skills which they failed to exhibit in the quizzes. Whether they gained those skills which they had initially failed to learn was tested by determining if they could exhibit the relevant skills in the Final Test. It would also have been possible to test for the effect of remediation by a further in-class quiz before instruction proceeded to the next skill. This was not done because it was felt to infringe too much upon class time and also because of the likelihood of spurious results from short-term memory. The actual sequence of events relating to instruction, remediation and testing is represented in Figure 8 (p. 126). The developmental and chemistry tests are described in the section which follows. The actual tests are presented in Appendices 1-6, while the two remedial units are presented in Appendices 8 and 9.

Instruments

Six test instruments were administered during the course of the study. Two developmental tests were applied, a 'Test of Developmental Level' consisting of three neo-Piagetian tasks and the Skemp test (Skemp, 1960) which measures the application of second-order operations. Four chemistry tests were applied, namely Quizzes One, Two and Three and the Final Chemistry Test. The composition of each will now be described briefly. Each is presented in detail in Appendices 1-6.

Test of Developmental Level

A battery of four tasks each derived from the Piagetian model of development was involved.

The first task tests the subject's concept of conservation of displaced volume. To answer correctly the subject must realize that the volume of a liquid displaced by a solid object is determined solely by the volume of the object. Hence he should predict that the water level in a container will be displaced to the same level when two objects of different weight but the same volume are immersed separately in the same volume of water in identical containers. Renner and Stafford (1972, p. 293) suggest that this task is a useful indicator of early formal operational thought. Piaget, Inhelder and Szeminska (1960, p. 382) indicate that understanding that interior volume is identical to displaced volume indicates the availability of early formal operational thought. In the present study two points were awarded for a correct prediction and explanation, one point for an incorrect prediction followed by a correct explanation after the result was observed, and no points for an incorrect explanation whether the prediction was correct or not.

The second task, represented as the 'Radio Task,' tests the subject's facility in combining variables. It was developed by Hobbs (1975) as a modification of a task devised by Bredderman (1974). In each case the subject is required to produce as many combinations as he can from a restricted set of objects and in each case the maximum possible number of combinations is 16. Hobbs judged an individual to be formal only if he was able to identify all 16 combinations, a procedure that is not consistent with Piaget's interpretation of the related chemical combinations task (Inhelder and Piaget, 1958). Conversely, Bredderman is consistent with Piaget in that he was concerned to identify a systematic approach rather than the complete number of combinations. In the present study Hobbs' task was used for its ease of application by the investigator in a large group setting. However, a subject was judged formal and given two points for systematically obtaining at least 13 combinations, one point for systematically obtaining between 9 and 12 combinations, and no points for any number of combinations arrived at unsystematically or for obtaining less than 9 combinations. As well as being a useful indicator of formal thought in general the task may also be considered appropriate in the present study in the sense that successful completion of the question testing the upper skills in the hypothesized hierarchy involves systematic selection and combination of lower skills.

The third task represents a minor modification by Wheeler (1976) of a task developed by Karplus and Karplus (1970) which was described in Chapter four. The task tests the ability to apply ratio in a novel situation, an ability clearly related to the content of the hypothesized hierarchy as well as being a more general indicator of developmental level. Subjects who reacted to the task by applying a ratio strategy were considered at least early formal and were awarded two points, whether or not their answer was perfectly correct. Subjects who applied an additive strategy were awarded one point, while those who did not apply either a ratio or an additive strategy were awarded a score of zero.

A fourth task was derived from Inhelder and Piaget's (1958) task on reflection. However subjects' responses to this task were uninterpretable because many of them were described only in terms of the subject's empirical experiences of playing pool.

The group of four tasks took a maximum of 40 minutes to administer, but no actual time limit was imposed. The protocol for administration is represented in Appendix 7.

The Skemp Test

A copy of this test is given in Appendix 6. It may also be considered to be a test of an individual's intellectual development, especially in the sense that correct responses to the component items appear to depend mainly on the ability to apply second-order operations. Piaget (1958) has referred to this as an aspect of formal operational thought. Lunzer (1966) stresses second-order operation as the essential nature of formal thought, although later he de-emphasizes this aspect to some extent (1973).

The relationship between Gagnéan and Piagetian theory in this regard is intriguing. It appears that Gagné's higher order rules and Piaget's second-order operations may be quite congruent. It would not, therefore, be surprising if difficulties with the upper levels of Gagnéan hierarchies were paralleled by an inability to display formal operational thought, at least as measured by an instrument testing the ability to perform second-order operations.

Administration of the Skemp test takes the form of practice at identifying and applying 10 basic operations, followed by three fifteen-item sub-tests requiring the reversal, combination, and reversal plus combination, respectively, of selections of the original ten operations. The test was devised by Skemp (1961) and is used with his permission in accordance with his instructions. It has been used in previous studies at the University of Alberta by Harrison (1967), Kass (1969), and Wheeler (1973). Harrison's modification is the one actually used.

Final Chemistry Test

This is a thirty-three item test including eight multiple-choice and 25 free-response items designed by the investigator to test mastery of the content represented by the hypothesized hierarchy. Each skill of the hierarchy is represented by four questions, two for chemical elements and two parallel questions for compounds. The order of questions is scrambled to prevent bias in favor of the hierarchy. One additional item consisting of six parts tests the ability to calculate molecular weights of compounds. It was intended that subjects who performed poorly on this question and who also failed to exhibit mastery of Skill E for compounds would be deleted from the sample when considering the general hierarchy in which discrimination was not made between performance of the same skill for elements and compounds. No such cases were found.

The test was expected to take about one hour. However, it is essentially compounded from a number of smaller criterion-referenced tests. In a criterion-referenced test it is important that all subjects should have adequate time to attempt all questions. Teachers were advised of the importance of this and were instructed to ensure it. In no case did the test take more than 70 minutes.

Chemistry Quizzes

Three quizzes were produced, embodying parallel items to those in the Final Test. Quiz One contains eight multiple-choice items, four testing Skill H (atomic and molecular weight as a ratio) and four testing Skill G (mole in terms of atomic and molecular weight and the Avagadro number of particles). Quiz Two tests understanding of Skills C (mole: relative # particles), D (mole: # particles), E (mass: equal

moles) and F (mass: # moles). Each skill is tested by two items relating to chemical elements and two parallel items relating to chemical compounds. Items testing Skills D and F contain two parts which were considered as two items for the purpose of the White and Clark test. Quiz Three tests understanding of Skill A (mass: relative # particles) and Skill B (mass: # particles). Each skill is tested by two items relating to chemical elements and two parallel items relating to chemical compounds. Collectively the three quizzes represent a parallel test to the Final Test. No time limit was imposed for any of the quizzes. None needed a full class period for completion by all subjects.

About a week before the study commenced each of the quizzes and the Final Test were field-tested on two groups of students enrolled in the same introductory chemistry course as the subjects in the study. One group was administered the quizzes and the other the Final Test. Each of these two groups was taught by the same teacher in one of the participating schools, but the teacher was not involved in the main study. The two groups were chosen because they had just completed their treatment of the mole concept. The teacher and students were asked to comment on the test items and the length of the tests. All students were able to complete either the group of three quizzes or the Final Test within one double period. Minor modifications were made to some of the items as a result of comments by the teacher and students involved.

The content validity of the Final Test was determined by asking three independent judges to rate each test item according to its congruence with the skill involved. Complete agreement was observed. The same procedure was not followed for the quizzes because the items contained in these were parallel to the items in the Final Test.

The problem of determination of reliability of criterion-referenced tests was discussed in Chapter two. A completely satisfactory test does not appear to be available. It is proposed that a focus on parallel internal consistency is most appropriate. Certainly questions representing the same skill should correlate significantly, although this might be considered a minimum criterion. Swaminathan, Hambleton and Algina's (1974) test, discussed in Chapter two, seems to be the most appropriate. The results of applying this test are reported in the next chapter.

Chapter 6

RESULTS AND DISCUSSION

Introduction

The basic data used to test the validity of an hypothesized learning hierarchy are derived from the responses of a number of individuals to tests of the intellectual skills represented in the hierarchy. The validity and reliability of these tests are therefore of much importance. Chapter 6 begins with a consideration of these parameters for the tests applied in the present study. This is followed by application of White and Clark's (1973) test of inclusion, the ordering-theoretic method developed by Bart and others (Bart and Krus, 1973; Airasian and Bart, 1975) and the Dayton and Macready (1976) scaling method to the data from the tests of the skills in the hypothesized hierarchy. The results of these analyses allow for meaningful comparisons of the application of three different methods to the same data as well as allowing a judgement of the validity of the hypothesized hierarchy and possible modifications of this hierarchy. As a result of this a 'preferred' hierarchy is suggested which is then further tested by considering the degree of transfer of learning from subordinate skills to related superordinate skills. The hierarchy remaining at this stage is considered to be the hierarchy validated by the present study.

A further aspect of the study refers to the relationship between learner developmental level and acquisition of the skills in the validated hierarchy. The strength of this relationship is determined by application of correlation and regression analyses. Finally, the relative efficiency of scores on the tests for subordinate skills and the developmental test scores as predictors of achievement on the tests for related superordinate skills in the validated hierarchy is considered.

For the application of the Dayton and Macready method the computer programme developed by Dayton and Macready (1976b) was used. For the application of the White and Clark test a specially written computer programme (Cornish, R., 1978) was used. All other statistical procedures were performed using the SPSS 300 statistical package (Nie et al., 1975). In all cases an IBM 370 (model 158) computer located at the Memorial University of Newfoundland was used.

Validity of the Chemistry Skills Tests

The ordering-theoretic method and the Dayton and Macready methods as well as the test of transfer used in the present study require a decision of the mastery status of each individual with respect to each of the intellectual skills in the hierarchy under test. In addition the White and Clark test, although not requiring a mastery decision, requires that the items testing a particular skill should exhibit low inter-item variance. Hence the three quizzes and the Final Test are each essentially composed of several criterion-referenced tests, one for each skill being tested.

Gagné (1969b) suggests that the most important consideration in the construction of criterion-referenced tests is that they should have good content validity. To ensure this three experienced chemistry teachers were asked to rate the congruence between each test item and the corresponding behavioral statement of the related skill. No items were considered incongruent by all three teachers. In the few cases where an individual teacher found an item to be incongruent the investigator discussed the item with the individual teacher. Following this procedure all items were considered to be acceptable by all three teachers.

Despite the agreement indicated above a further test was applied to the test items before consideration of the reliability of the chemistry skills tests. The test applied was to determine the phi coefficient as an index of the degree of correlation between each item testing the same skill in each particular test. Ideally, any individual should either answer all items testing a particular skill correctly or incorrectly. In practice such perfect agreement is seldom found, especially as items may usually be considered to be a sample from a particular domain rather than being identical items. Moreover, the values of phi coefficients are affected by the marginal totals of the contingency tables from which they are determined. Accordingly, phi coefficients significantly greater than zero were considered necessary between each item testing a particular skill, but perfect correlations were not anticipated. The values obtained are presented in Tables 1 to 8. As well as representing the strength of the relationships involved these matrices may be used to determine if any items are particularly deviant from the others testing the same skill. The value of N varies between tests partly because some subjects were absent on a particular occasion and partly because subjects whose responses to any particular item were difficult to interpret

Table 1

Phi Coefficients Between Items Testing Skill A
(mass: relative # particles)

Quiz Three

Item	303	304	307	308
303	--			
304	.66	--		
307	.61	.61	--	
308	.49	.45	.54	--

N = 205

ϕ = 0.23 significant at 0.001 level.

Final

Item	19	23	29	32
19	--			
23	.50	--		
29	.45	.48	--	
32	.43	.43	.45	--

N = 239

ϕ = 0.21 significant at 0.001 level.

Note: In the above and subsequent matrices relating to quizzes the first digit for each item refers to the quiz number and the remaining two digits to the number of the item in the quiz. For the Final Test each item is represented by the number of the item in that test.

Table 2

Phi Coefficients Between Items Testing Skill B
(mass: # particles)

Quiz Three

Item	301	302	305	306
301	--			
302	.51	--		
305	.52	.51	--	
306	.45	.50	.36	--

N = 212

ϕ = 0.23 significant at 0.001 level.

Final

Item	02	10	17	24
02	--			
10	.44	--		
17	.38	.46	--	
24	.45	.48	.69	--

N = 248

ϕ = 0.21 significant at 0.001 level.

Table 3

Phi Coefficients Between Items Testing Skill C
(mole: relative # particles)

Quiz Two

Item	209	210	211	212
209	--			
210	.59	--		
211	.59	.73	--	
212	.57	.90	.78	--

N = 213

ϕ = 0.23 significant at 0.001 level.

Final

Item	04	08	14	28
04	--			
08	.52	--		
14	.58	.55	--	
28	.73	.49	.58	--

N = 249

ϕ = 0.21 significant at 0.001 level.

Table 4

Phi Coefficients Between Items Testing Skill D
(mole: # particles)

Quiz Two

Item	201a	201b	202a	202b	203a	203b	204a	204b
201a	--							
201b	.92	--						
202a	.53	.50	--					
202b	.55	.50	.82	--				
203a	.58	.54	.60	.67	--			
203b	.56	.53	.65	.68	.86	--		
204a	.40	.38	.51	.46	.57	.53	--	
204b	.44	.44	.70	.68	.59	.60	.50	--

N = 251

ϕ = 0.21 significant at 0.001 level.

Final

Item	03a	03b	07a	07b	09a	09b	21a	21b
03a	--							
03b	.64	--						
07a	.47	.50	--					
07b	.48	.49	.96	--				
09a	.40	.47	.63	.59	--			
09b	.42	.50	.61	.64	.89	--		
21a	.50	.62	.51	.51	.49	.51	--	
21b	.43	.40	.54	.52	.44	.40	.58	--

N = 233

ϕ = 0.22 significant at 0.001 level.

Table 5

Phi Coefficients Between Items Testing Skill E
(mass: equal # moles)

Quiz Two

Item	213	214	215	216
213	--			
214	.42	--		
215	.76	.49	--	
216	.53	.54	.59	--

N = 207

ϕ = 0.23 significant at 0.001 level.

Final

Item	01	15	25	31
01	--			
15	.19*	--		
25	.15**	.50	--	
31	.21	.22	.33	--

N = 246

ϕ = 0.21 significant at 0.001 level

*significant at 0.01 level

**significant at 0.05 level

Table 6

Phi Coefficients Between Items Testing Skill F
(mass: # moles)

Quiz Two

Item	205a	205b	206a	206b	207a	207b	208a	208b
205a	--							
205b	.72	--						
206a	.52	.42	--					
206b	.55	.45	.95	--				
207a	.55	.46	.46	.49	--			
207b	.43	.43	.44	.47	.74	--		
208a	.41	.48	.53	.56	.44	.55	--	
208b	.41	.49	.43	.47	.47	.51	.55	--

N = 248

ϕ = 0.21 significant at 0.001 level.

Final

Item	18a	18b	22a	22b	27a	27b	30a	30b
18a	--							
18b	.57	--						
22a	.29*	.27	--					
22b	.26	.36	.54	--				
27a	.36	.29	.15**	.31	--			
27b	.25	.29	.18*	.27	.60	--		
30a	.17*	.33	.22	.27	.30	.27	--	
30b	.26	.30	.25	.39	.33	.32	.72	--

N = 233

ϕ = 0.22 significant at 0.001 level.

*significant at .01 level.

**significant at .05 level.

Table 7

Phi Coefficients Between Items Testing Skill G
(mole: molar mass, L particles)

Quiz One

Item	103	104	107	108
103	--			
104	.34	--		
107	.26	.74	--	
108	.71	.31	.34	--

N = 254

ϕ = 0.21 significant at 0.001 level.

Final

Item	05	13	20	33
05	--			
13	.74	--		
20	.29	.31	--	
33	.25	.23	.54	--

N = 254

ϕ = 0.21 significant at 0.001 level.

Table 8

Phi Coefficients Between Items Testing Skill H
(molar mass as a ratio)

Quiz One

Item	101	102	105	106
101	--			
102	.00	--		
105	.47	-.05	--	
106	-.01	.35	-.12	--

N = 255

ϕ = 0.21 significant at 0.001 level.

Final

Item	06	11	16	26
06	--			
11	.05	--		
16	.18	.02	--	
26	.07	.31	.11	--

N = 254

ϕ = 0.21 significant at 0.001 level.

were treated as having missed the test containing that item.

The values of the phi coefficients reported in Tables 1 to 8 are almost all significantly different from zero at the 0.001 level of confidence. However, several anomalies are apparent and these will now be discussed.

As may be seen in Tables 5 and 6, for Skills E (mass: equal # moles) and F (mass: # moles) the phi coefficients obtained between items in the Final Test are generally considerably lower than those between parallel items in Quiz Two. One interpretation of this is that several items in the Final Test sub-tests for Skills E and F do not represent the same skill as the remaining items. In particular, Item 01 testing Skill E and Items 18a and 22a testing Skill F appear to be deviant. However, further reference to Table 5 indicates that Item 214 in Quiz Two which is parallel to Item 01 in the Final Test correlates significantly (0.42, 0.49, 0.54) with each of the other items testing Skill E in Quiz Two. Similarly, the phi coefficients between Items 207a and 208a (which are parallel to Items 18a and 22a, respectively) with the other items testing Skill F in Quiz Two range from 0.41 to 0.74 and are also considered to be highly significant. Hence, it does not appear likely that Items 01, 18a and 18b are deviant in the sense that they are unrepresentative of the skills which they are intended to represent. Therefore, no items were eliminated from the tests for Skills E and F.

A different interpretation appears likely for several extremely low phi coefficients between items testing Skill H (molar mass as a ratio), which are represented in Table 8. The items in this sub-test separate into two distinct pairs in which significant phi coefficients are observed between items within each pair and non-significant phi coefficients are observed between any item in either pair and any item in the other pair. Hence, consideration was given to the subdivision of Skill H into two sub-skills, each tested by a pair of items. One of these potential sub-skills, tested by Items 102 and 106 in Quiz One and by Items 06 and 16 in the Final Test may be considered to require discrimination of formalized definitions of atomic and molecular weights from several plausible alternatives. This sub-skill was rejected after further consideration on the grounds that it required recall of verbalized knowledge rather than an intellectual skill. The other potential sub-skill, tested by Items 101 and 105 in Quiz One and by Items 11 and 26 in the Final Test, focuses upon relative atomic and molecular weights of substances as the ratio of the weights of the atoms and molecules involved rather than upon discrimination of a formalized definition. This was retained as a modified form of Skill H. Further references to Skill H relate to this aspect of the original skill.

Reliability of the Chemistry Skills Tests

The determination of the reliability of each of the tests for chemistry skills requires first a consideration of the most appropriate type(s) of reliability. In the present study each test for chemistry skills is used primarily to determine the mastery or non-mastery status of each individual subject with respect to the particular skill. Hence, the most appropriate estimate of reliability should relate to the consistency of mastery/non-mastery decisions. An appropriate measure of this is Cohen's κ (Cohen, 1960). This was applied to determine the consistency of mastery decisions for each subject with respect to each skill when tested in the quizzes and Final Test respectively.

To avoid distortion of κ by gain of skills due to remediation, only the data from those subjects who were not administered the remedial units were included in this analysis. Each of these subjects was compared according to mastery or non-mastery status for each skill as determined by the appropriate quiz and Final Test respectively, and the value of κ for this sub-sample then determined. To determine if the coefficients obtained were significantly different from zero the coefficients were converted to standard scores and the significance of these scores determined by reference to the normal curve. The results are indicated in Table 9.

With the exception of the coefficient for Skill F all of the estimates of κ are significantly greater than zero at the 0.001 level of confidence. These results are particularly supportive of the consistency of mastery decisions based upon the tests when the data in Table 10 are also considered. Table 10 indicates the means and standard deviations for the complete non-remedial sub-group on the tests of the chemistry skills both in the quizzes and the Final Test. With the exception of Skill H, the mean for each skill was greater in the Final Test than in the corresponding quiz, hence tending to artificially decrease the observed value of κ.

Further, the apparently low level of agreement observed for Skill F may be explained in terms of the high mean scores on the tests involved. As indicated in Chapter 2 (p. 75), the expression derived by Cohen to estimate the measure of agreement between the two test

Table 9

Cohen's κ as a Measure of Test-Retest Reliability for Skills A to H

Element	κ	Z	Significance
A	.40	4.40	.001
B	.56	5.33	.001
C	.64	5.87	.001
D	.64	7.20	.001
E	.42	4.04	.001
F	.35	2.05	N.S.
G	.44	5.51	.001
H	.25	3.38	.001

Table 10

Means and Standard Deviations for the Non-Remedial Sub-Group on the Tests of Skills A to H in the Quizzes and Final Test

Skill	N	Quiz Mean	Quiz S.D.	Final Mean	Final S.D.
A	101	3.2	3.3	3.8	3.1
B	112	2.8	3.1	4.1	3.2
C	116	2.7	3.3	3.1	3.3
D	143	3.0	3.2	4.2	3.2
E	112	4.6	3.2	6.1	2.3
F	136	6.8	2.2	7.3	1.4
G	143	3.7	3.1	5.3	2.6
H	151	7.3	1.8	7.0	2.1

Note: All scores scaled where necessary to give a maximum possible total score of eight.

administrations is

$$\kappa = (P_o - P_c) \ / \ (1-P_c)$$

while the significance of the value so obtained is determined by application of the expression

$$Z = \frac{\kappa}{\sqrt{\dfrac{P_c}{n(1-P_c)}}}$$

followed by reference of the Z value to the normal distribution. Simple arithmetic illustrates that when a large proportion of the subjects is judged to show mastery of the skill involved in both tests then the value of $P_o - P_c$ is small, $(1-P_c)$ is small, and hence κ is small. Hence an underestimate of κ is likely. This is compounded when the value of Z is determined as its value is also decreased when P_c is large and $(1-P_c)$ is small. The low value of κ for Skill F appears to be related to these factors rather than to a lack of agreement between the mastery decisions made on the two occasions.

In addition to their use as determinants of the mastery status of individuals with respect to the chemistry skills involved in the present study the chemistry skills tests were used to provide continuous scores representing degree of attainment of these skills. These scores were used in regression analyses performed to provide a test of hypotheses three and four. The stability of these continuous scores over time was estimated by determining Pearson product-moment correlations (r) between the scores on the parallel tests for each skill in the appropriate quiz and the Final Test respectively. In this analysis only the data from the non-remedial group were used, so that confounding through gains attributable to the effect of the remedial treatment would be eliminated. In addition one final estimate of reliability which was determined was Cronbach's alpha (α), which is equivalent to the Kuder-Richardson Formula 20 estimate of internal consistency. The need to provide an estimate of internal consistency derives from the use of White and Clark's test of inclusion as part of the hierarchical analysis. The White and Clark test assumes equivalence of items testing the same skill.

Before presenting the estimates obtained for r and α the use of each of these measures merits some discussion. Both depend upon variance in the individual scores involved. However, in the ideal case in a criterion-referenced test the variance should be small because subjects should either answer all test items from the same domain correctly or incorrectly. Hence an underestimate of reliability is likely when the product-moment coefficient or Cronbach's alpha is determined. However, in practice the variance is often sufficiently large to allow the meaningful use of these statistics. As is indicated in Table 10 (p. 151) this is true of the tests in the present study. The estimates determined for r and α are presented in Tables 11 and 12, respectively.

The values of the parallel forms estimates of reliability and the coefficients of internal consistency are generally sufficiently high to offer good support for the use of the tests applied in the present study to determine individual achievement of the chemistry skills involved. Once again, the most deviant values are those obtained for Skills F and H. Again the explanation for the relatively low values obtained for the tests of these skills appears to lie in the restricted range of scores on these tests.

Table 11

Parallel Forms Estimates of Test Reliabilities
(Non-Remedial Group)

Skill	Correlation	N	Significance
A	0.61	101	< .001
B	0.74	112	< .001
C	0.72	116	< .001
D	0.72	143	< .001
E	0.57	112	< .001
F	0.39	136	< .001
G	0.56	143	< .001
H	0.36	151	< .001

Table 12

Cronbach's Alpha as an Estimate of Internal Consistency for the Tests for Skills A to H

Quizzes

Skill	N	alpha
A	205	0.84
B	212	0.78
C	213	0.90
D	251	0.92
E	207	0.89
F	248	0.89
G	254	0.77
H	255	0.36

Final Test

Skill	N	alpha
A	239	0.77
B	248	0.79
C	249	0.84
D	233	0.90
E	246	0.85
F	233	0.79
G	245	0.72
H	254	0.33

In summary of the present section it is suggested that the use of the chemistry skills tests to provide decisions of the mastery status of individuals as well as continuous achievement scores may be approached with reasonable confidence. In the sections which follow the data from these tests will be used to answer the questions posed in Chapter 5.

Application of the White and Clark Test to the Data from the Final Test

As was indicated in Chapter 2 the White and Clark test appears to be the most appropriate test available at the present time to determine the existence of hierarchical connections between pairs of intellectual skills. Compared to the ordering-theoretic method the White and Clark test makes use of more information, does not require a mastery decision and provides a more rigorous statistical test. Consequently the White and Clark test in the modified form described in Chapter 2 was used as the primary method for determination of hierarchical connections between the skills under consideration in the present study when the skills are considered in pairs. In the discussion which follows wherever mention is made of the White and Clark test the modified version developed by Griffiths and Cornish (1978) is implied.

Research question one asks "*Does the arrangement of intellectual skills represented in the hypothesized hierarchy represent a learning hierarchy which is valid according to the application of the White and Clark test?*" If the answer to this question is negative Research question two will be considered. Research question two asks "*Does an arrangement of a sub-set of the skills represented in the hypothesized* *hierarchy exist such that the arrangement represents a learning hierarchy which is valid according to the White and Clark test?"*

Before considering these questions it is appropriate to first determine whether the appropriate level of analysis is represented by the skills in the hypothesized hierarchy (p.116) or by the subdivisions of these skills represented in the modification of the hypothesized hierarchy (p.118). The possibility of an internal hierarchical connection between sub-skills representing each skill in the hypothesized hierarchy was therefore considered. For each pair of sub-skills a sub-skill by sub-skill matrix to which the White and Clark test could be applied was constructed. The number of entries in the critical cell(s) for each possible internal connection in each direction is presented in Table 13. Skill H (molar mass as a ratio) is shown in parentheses because each sub-skill for Skill H was tested only by one question.

If an hierarchical relationship exists between the sub-skills comprising any particular skill the expected direction is such that the sub-skill representing chemical compounds should always be superordinate. The data represented in Table 13 support this for Skill D (moles: # particles) at the 00 level and for Skill F (mass: # moles) at the 01 level when the White and Clark test is applied. No hierarchical relationship is found within Skill A (mass: relative # particles), Skill B (mass: # particles), Skill E (mass: equal # moles) and Skill G (mole: molar mass, L particles). Surprisingly, for Skill C (moles: relative # particles) the direction of the connection is the inverse of what can be logically expected. The general pattern of these results denies the superiority of the modification of the hypo-

Table 13

White and Clark Test for Sub-Skills (Elements v. Compounds) within Major Skills (Final Test)

Skill	Number of Exceptions: Element / Compound	Number of Exceptions: Compound / Element	N
A	5	7	239
B	3	4	248
C	0	6	249
D	19	2	254
E	6	5	246
F	10	2	242
G	1	4	245
H	5	1	254

thesized hierarchy over the hypothesized hierarchy itself. Tentatively, an hierarchical relationship may be held to exist for the subskills within Skills D and F. However, retention of these 'logical' connections should be accompanied by retention of the 'illogical' connection within Skill C, in which exhibition of this skill with respect to chemical compounds appears to be subordinate to its exhibition for chemical elements. In view of the foregoing findings and considerations subdivision of the skills represented in the hypothesized hierarchy was not considered appropriate. In the analysis and discussions which follow reference to 'the hypothesized hierarchy' should be taken to mean the hierarchy originally hypothesized. Research questions one and two are now considered.

The results of applying the White and Clark test to the data from the Final Test are represented in Table 14. The test was applied to all pairs of skills where the data suggested the possibility of an hierarchical connection whether or not the connection was hypothesized originally. Hence in the table the designation 'upper' and 'lower' does not reflect any theoretical position. Each cell in the table contains the number of exceptions to the particular hierarchical connection (that is, the total number of entries in the critical cells), and in parentheses the level at which the connection is considered valid. With regard to this, three levels only are considered. These allow for entries which can be ascribed to errors of measurement only (00), errors of measurement plus one percent exceptions (01), and errors of measurement plus two percent exceptions (02). In all cases a five percent level of significance is used. Where a connection is not validated at any of the three levels, only the number of excep-

Table 14

White and Clark Test at Three Levels of Stringency: Number of Exceptions and Level at which Connection is valid

'Upper' Skill

$238 \leq N \leq 254$

		A	B	C	D	E	F	G	H
	A	--	16	4 (00)	12	45	73	27	73
	B	21	--	7 (02)	3 (00)	57	82	28	84
	C	27	26	--	19	79	106	46	107
'Lower' Skill	D	14	4 (01)	1 (01)	--	43	61	18	60
	E	0 (00)	4 (01)	1 (01)	3 (01)	--	16	7 (02)	18
	F	0 (00)	0 (00)	0 (00)	0 (00)	0 (00)	--	0 (00)	1 (01)
	G	8 (02)	8 (02)	2 (01)	3 (00)	23	35	--	33
	H	3 (01)	3 (01)	0 (00)	2 (00)	4 (01)	6 (02)	1 (00)	--

tions is recorded. In terms of the hierarchical relationships which emerge, the results in Table 14 may be more informatively expressed as in Table 15. The latter indicates those skills judged to be subordinate to each of the other seven skills, at each of the three levels of stringency applied.

In determining the hierarchies that emerge from the data only those hierarchies leading to Skill A as superordinate were considered, since this was the desired capability in this study. Depending on the stringency applied to the White and Clark test, three alternative hierarchies emerge. In order of decreasing stringency these are represented in Figures 9, 10 and 11.

Consideration of these hierarchies raises several questions. Firstly, given three hierarchies to choose from, which is most appropriate? It would appear that the more stringent the test the more certain one can be of the validity of the hierarchy. However, this is not necessarily so. Some connections appear to be only marginally established at one level because they disappear at a less stringent level, the relationship becoming bi-directional. For example, Skill H (molar mass as a ratio) is not found in Hierarchy One, appears to be superordinate to Skill F (mass: # moles) in Hierarchy Two, and in Hierarchy Three is equivalent to Skill F. A similar relationship is found for B and D, although these skills do not feature in Hierarchies One to Three. In general it may be suggested that provided all skills are of use in themselves larger hierarchies are more useful guides than smaller ones to the arrangement of instruction because they contain more information. However, while this might suggest testing at much less stringent levels such a procedure would not be particularly

Table 15

Summary of Hierarchical Connections Identified
After Application of the White and Clark Test

Level	A	B	C	D	E	F	G	H
00	E F	F	A F H	B F G H	F		F H	
01	E F H	D E F H	A D E F G H	B E F G H	F H		F H	F
02	E F G H	D E F G H	A B D E F G H	B E F G H	F H	H	E F H	F

Note: The reader's interpretation of the above and similar tables may be aided by an illustrative example. Table 15 indicates that Skill A is superordinate to Skills E and F at the 00 level; to Skills E, F and H at the 01 level; and to Skills E, F, G, and H at the 02 level.

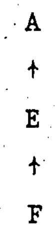


Figure 9. Hierarchy One from application of the White and Clark Test at the 00 Level to the Final Test Data.

Figure 10. Hierarchy Two from application of the White and Clark Test at the 01 Level to the Final Test Data.

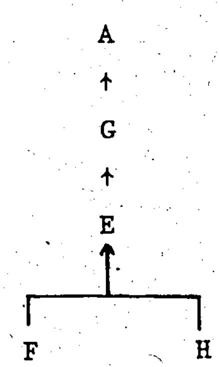


Figure 11. Hierarchy Three from application of the White and Clark Test at the 02 Level to the Final Test Data.

useful because, taken to the ultimate extreme, the relationship between the skills in each pair would be validated in both directions. Hence there would be no hierarchy. At the opposite extreme only exceptions attributable to error of measurement would be allowed. This represents the absolute notion of an hierarchy, but may result in very small hierarchies of little practical use. As suggested by Gagné (1970b) substantial rather than absolute levels of hierarchical dependence seem more appropriate, but what constitutes an optimum level is less clear. The present investigator suggests, in the absence of better criteria, that it is appropriate to determine the hierarchies which emerge under progressively less restrictive criteria, indicating the hierarchies and the allowed percentage of exceptions in each case. When the number of uni-directional hierarchical connections begins to decrease analysis should cease. In the present study this occurs when analysis is continued beyond the 02 level.

In the hierarchies considered above some of the connections may be considered tenuous. In Hierarchy Two, Skill H is superordinate to Skill F, whereas in Hierarchy Three they are equivalent. The difference between them with respect to their empirical relationship to the other skills is not substantial. Hence Hierarchy Four, represented in Figure 12, is considered a reasonable alternative to Hierarchy Two.

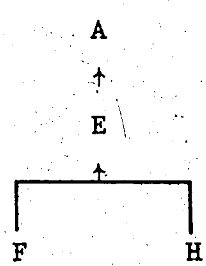
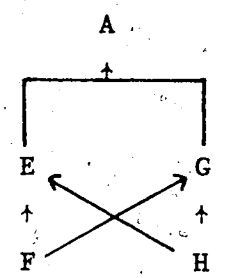
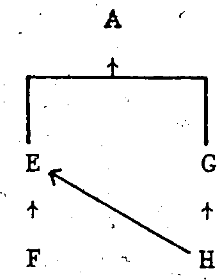


Figure 12. Hierarchy Four as an alternative to Hierarchy Two.

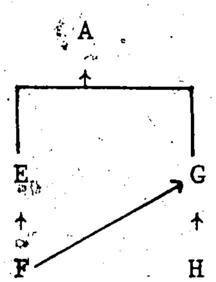
A second potential anomaly relates to the position of Skill G (mole: molar mass, L particles) which is not empirically subordinate to Skill A (mass: relative # particles) until the 02 level. In addition it is difficult to justify the superordinate position of Skill G relative to Skill E (mass: equal # moles) and Skill F (mass: # moles) in terms of the logical relationships between Skill G and Skills E and F respectively. White (1974a) suggests that connections which are illogical in terms of the subject matter concerned should be ignored. This position is difficult to accept for two reasons. First, the logic of the subject matter is not necessarily congruent with the psychological structure of the learner. Second, it does not seem reasonable to allow the researcher to accept some connections and to reject others when the empirical evidence with respect to both may be the same. Indeed, if White's contention is accepted the appropriate method for generating learning hierarchies would be by logical introspection by experts, unaccompanied by empirical validation. The history of learning hierarchy research suggests such practice to be ineffective. Nevertheless, further consideration of Hierarchy Three in the light of the illogical relationship between Skill G and Skills E and F, together with the marginal connection between Skills E and H suggests the need to investigate other alternative arrangements of the skills represented in Hierarchy Three. Hierarchies Five to Eight which are represented in Figure 13 are suggested as possible alternatives to Hierarchy Three. Fortunately, the Dayton and Macready method allows a test of the relative appropriateness of alternative hierarchies derived from the same data. This is investigated further in a later section.



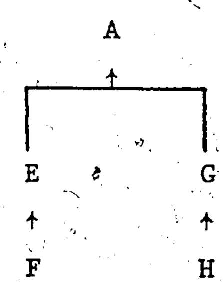
Hierarchy Five



Hierarchy Six



Hierarchy Seven



Hierarchy Eight

Figure 13. Hierarchies Five to Eight as alternatives to Hierarchy Three.

Research question one has been answered negatively. Application of the White and Clark test to the data indicates that the hypothesized hierarchy is not supported in its entirety. Research question two has been answered positively. On the basis of the White and Clark test it is possible to propose an alternative hierarchy (Hierarchy One) which is composed from a sub-set of the skills represented in the hypothesized hierarchy. If Linke's modification to allow for one and two percent exceptions to each hierarchical connection is adopted, Hierarchy Two and Hierarchy Three, respectively, may be considered valid. It is further suggested here that because Hierarchy Three contains more skills than Hierarchies One and Two it is the most useful of the three to the designer of instruction. As well as these three hierarchies five others (Hierarchies Four to Eight inclusive) have been suggested as possible alternatives following consideration of the connections involved. Before considering the validity of these hierarchies by application of the Dayton and Macready method, another method which compares skills in pairs, the ordering-theoretic method, will be applied to the same data as the White and Clark test. This is the focus of the next section.

Application of the Ordering-Theoretic Method

The greater precision of the White and Clark test compared to the ordering-theoretic method has already been noted in Chapter 2. However, the recent popularity of the ordering-theoretic method suggests a comparison of the results obtained from application of each to be appropriate. Moreover, the ordering-theoretic method is easier to use and if it produces similar results to the White and Clark test it may be used in preference to the White and Clark test. The relationship between the two methods is the focus of research question three, which asks, "*What is the extent of agreement between the results of applying the White and Clark test and the ordering-theoretic method to the same data?*"

The correct classification of subjects as masters or non-masters of a skill is a problem in this study, as in other similar ones. For the application of both the ordering-theoretic method and the Dayton and Macready method such classification is necessary. Therefore, each analysis has been performed twice, first using the complete sample and then the same sample restricted by eliminating subjects with central scores. Hence, in the first instance scores of at least one, three and six points were taken to represent possession and lesser scores to represent absence of a skill tested by two, four and eight items respectively. In the second instance scores of at least two, three and six points were taken to represent possession of the same skills while scores of less than one, two and three, respectively, represented lack of these skills; it was considered that subjects with intermediate scores could not be clearly classified as possessing or lacking the skills and these subjects were dropped from the second analysis. The intent of this second classification was to increase the certainty of classification, and hence to increase confidence in the validity of any hierarchy which may emerge. A second possibility is that different hierarchies may exist if one considers partial possession of skills as opposed to possession and non-possession.

Hierarchical connections were tested by the ordering-theoretic method for both the restricted and unrestricted sample, allowing successively for one, two and five percent exceptions. The results are indicated in Tables 16 and 17. As in the case of the White and

Table 16

Ordering-Theoretic Method: Percentage of Exceptions to Hierarchical Connections (Complete Sample, Final Test)

'Lower' Skill	'Upper' Skill							
	A	B	C	D	E	F	G	H
A	--	11	4.5	8.7	28	42	18	42
B	14	--	2.8	7.4	35	46	21	45
C	18	17	--	7.0	42	56	26	55
D	11	6.9	2.0	--	18	46	17	49
E	1.3	4.0	1.2	3.7	--	18	6.5	19
F	0.0	0.4	0.0	0.4	3.0	--	1.6	7.5
G	8.0	7.1	2.0	4.0	24	33	--	32
H	1.2	1.2	0.4	0.8	1.6	2.4	0.4	--

Table 17

Ordering-Theoretic Method: Percentage of Exceptions to Hierarchical Connections (Restricted Sample, Final Test)

		'Upper' Skill							
		A	B	C	D	E	F	G	H
'Lower' Skill	A	--	8.9	2.2	7.2	25	39	17	40
	B	12	--	3.7	1.8	32	46	18	44
	C	15	14	--	12	44	56	29	55
	D	8.4	2.4	0.6	--	27	36	12	36
	E	0.0	2.2	0.6	1.9	--	8.2	4.4	9.6
	F	0.0	0.0	0.0	0.0	0.0	--	0.0	0.5
	G	5.0	5.0	1.3	2.0	15	20	--	21
	H	1.6	1.6	0.0	1.2	2.1	3.1	0.6	--

Clark test all possible connections were considered. Tables 18 and 19 summarize the connections identified from consideration of the results represented in Tables 16 and 17. To facilitate direct comparison of the results of applying the ordering-theoretic method and the White and Clark test to the data in the present study, Table 20 was compiled. It summarizes the information presented in Tables 16, 17, 18 and 19. Cohen's κ was used to determine the level of agreement between the results of applying the two methods. The results of this analysis are presented in Table 21.

Examination of Table 21 suggests that, irrespective of whether it is applied to the restricted sample or the whole sample, the result of applying the ordering-theoretic method at the 1% level is most similar to the result of applying the White and Clark test at the 00 level. At the 2% level the result of using the ordering-theoretic method is most similar to that for the White and Clark test at the 01 level. Finally, at the 5% level, the result of using the ordering-theoretic method is most similar to that for the White and Clark test at the 02 level. A further finding is that the results from the use of the White and Clark test are more consistent with those from use of the ordering-theoretic method for the restricted sample than with the unrestricted sample at each level of stringency. This better fit could be attributed to greater precision in classifying subjects, or alternatively to greater congruence between the implicit mastery criterion in the White and Clark test and the mastery criterion for the restricted sample when the ordering-theoretic method is applied relative to that used when the sample is unrestricted. The complete agreement between the results of applying the ordering-theoretic

Table 18

Summary of Hierarchical Connections Derived by Applying the Ordering-Theoretic Method (Complete Sample, Final Test)

Level	A	B	C	D	E	F	G	H
1%	F	F	F H	F H				
2%	E F H	F H	D E F G H	F H	F		F H	
5%	E F H	D E F H	A B D E F G H	E F G H	F H	H	F H	

Table 19

Summary of Hierarchical Connections Derived by Applying the Ordering-Theoretic Method (Restricted Sample, Final Test)

Level	A	B	C	D	E	F	G	H
1%	E F	F	D E F H	F	F		F H	F
2%	E F H	F H	D E F G H	B E F G H	F		G H	F
5%	E F G H	D E F G H	A B D E F G H	F E F G H	F H	H	E F H	F

Table 20

Hierarchical Connections Derived by Applying the Ordering-Theoretic and White and Clark Tests

Connection (first over second)	Ordering-Theoretic (whole range)			Ordering-Theoretic (restricted range)			White and Clark		
	1%	2%	5%	1%	2%	5%	00	01	02
A-B									
A-C									
A-D									
A-E		1	1	1	1	1	1	1	1
A-F	1	1	1	1	1	1	1	1	1
A-G						1			1
A-H		1	1		1	1		1	1
B-A									
B-C									
B-D			1			1		1	1
B-E			1			1		1	1
B-F	1	1	1	1	1	1	1	1	1
B-G						1			1
B-H		1	1		1	1		1	1
C-A			1			1	1	1	1
C-B			1			1			1
C-D		1	1	1	1	1	1	1	1
C-E		1	1	1	1	1		1	1
C-F	1	1	1	1	1	1	1	1	1
C-G		1	1		1	1		1	1
C-H	1	1	1	1	1	1	1	1	1

....continued

Note: 1 = 0 validated connection, blank = non-validated connection.

Table 20 (continued)

Connection (first over second)	Ordering-Theoretic (whole range)			Ordering-Theoretic (restricted range)			White and Clark		
	1%	2%	5%	1%	2%	5%	Q0	01	02
D-A									
D-B					1	1	1	1	1
D-C									
D-E			1		1	1		1	1
D-F	1	1	1	1	1	1	1	1	1
D-G			1		1	1		1	1
D-H	1	1	1		1	1	1	1	1
E-A									
E-B									
E-C									
E-D									
E-F		1	1	1	1	1	1	1	1
E-G									
E-H			1			1		1	1
F-A									
F-B									
F-C									
F-D									
F-E									
F-G									
F-H			1			1			1

....continued

Table 20 (continued)

Connection (first over second)	Ordering-Theoretic (whole range)			Ordering-Theoretic (restricted range)			White and Clark		
	1%	2%	5%	1%	2%	5%	00	01	02
G-A									
G-B									
G-C									
G-D									
G-E						1			1
G-F		1	1	1	1	1	1	1	1
G-H	1	1	1	1	1	1	1	1	1
H-A									
H-B									
H-C									
H-D									
H-E									
H-F				1	1	1		1	1
H-G									

Table 21

Cohen's κ as a Measure of Agreement Between the Ordering-Theoretic Method and the White and Clark Test

		White and Clark		
		00	01	02
Ordering-Theoretic - (unrestricted)	1%	.61	.31	.21
	2%	.66	.69	.53
	5%	.45	.85	.86
Ordering-Theoretic - (restricted)	1%	.68	.56	.42
	2%	.64	.85	.68
	5%	.42	.82	1.00

method to the restricted sample at the 5% level and the White and Clark test at the 02 level encourages the interpretation that results from application of these two tests may be used interchangeably. The measure of agreement observed in Table 21 for the comparisons at the other levels is similarly encouraging. However, the absolute number of differences observed when these tests are applied to the same data in the present study suggests the need for caution. Comparing the application of the ordering-theoretic method with the restricted sample to the application of White and Clark test, six, four, and zero different connections were observed at the three successive stringency levels. Using the unrestricted sample for the ordering-theoretic method in similar fashion six, seven, and four differences were observed. In the present study little or no difference would have been found in the hierarchies suggested at each level if the ordering-theoretic method, especially for the restricted sample, had been used in place of the White and Clark test. However, this may have been fortuitous because many of the deviant results related to Skill B (mass: # particles) which did not feature in the eventually preferred hierarchies. Nevertheless the answer to research question three is that the results, particularly at the least stringent level, are at least suggestive that the ordering-theoretic method may be substituted for the White and Clark test. Whichever of these two tests is used the fact remains that the resultant hierarchy is a composite and is not tested directly. The Dayton and Macready method tests the validity of an hierarchy as a complete entity. In the next section it is applied to determine the goodness of fit of Hierarchies One to Eight to the data.

Application of the Dayton and Macready Method

Research question four asks, "*Does the hypothesized hierarchy, or an alternate hierarchy composed of some or all of the skills contained in the hypothesized hierarchy, represent a learning hierarchy which is valid according to the application of the Dayton and Macready method?*" This question is addressed in the present section.

The Dayton and Macready method was described in detail in Chapter 2. Goodness of fit between data and an hypothesized hierarchy is determined by both a chi-square analysis and the determination of a likelihood ratio expressed as a chi-square. Both statistics suffer from the fact that for any hierarchy containing more than three skills the number of possible response patterns which are inconsistent with the hierarchy exceeds the number of response patterns which are consistent with it. Hence a large proportion of the observed frequencies should be zero if the hierarchy is valid. Unfortunately, the value of the chi-square statistic may therefore be substantially affected by as little as one response which is "false negative" (that is, when the subject falsely appears to lack a skill relatively low in the hierarchy while exhibiting one or more related higher skills). The value of the likelihood function is also affected adversely although it is more robust than the chi-square statistic. In the light of these comments correct classification of subjects as masters or non-masters of each component skill is particularly important. For this reason the Dayton and Macready method was applied using the same restricted sample as for the application of the ordering-theoretic method, as well as for the complete sample.

In computing the value of the likelihood ratio the Dayton and

Macready method yields estimates of the guessing and forgetting parameters needed to provide a fit between data and hierarchy. As these values increase, confidence in the particular hierarchy decreases. They may be used to aid differentiation between alternative hierarchies, although the primary statistic used for hierarchies containing an equal number of skills is the difference between likelihood ratios. Hierarchies One to Eight were therefore subjected to analysis by Dayton and Macready's method. Table 22 indicates the values of the misclassification parameters and the likelihood ratios for these eight hierarchies when the data from the complete sample were used without elimination of central scores. The hypothesized hierarchy was not subjected to similar analysis because the frequency of response patterns inconsistent with the validity of this hierarchy was clearly too high. The null hypothesis under test is that there is no significant difference between the observed frequencies and those expected if the hierarchy in question is considered true. For the unrestricted sample none of the three hierarchies (Hierarchies One, Two and Three) suggested after applying the White and Clark test was consistent with the data when the Dayton and Macready test was applied at the five percent level. However, Hierarchies Four and Seven, suggested as alternates to Hierarchies Two and Three, respectively, were consistent with the data when the same test was applied. Further, another alternate to Hierarchy Three (Hierarchy Eight) was only marginally inconsistent at the five percent level.

The question of whether Hierarchy Seven or Hierarchy Four is most appropriate is perhaps best answered on utilitarian grounds. In general it is suggested that given the choice of two hierarchies which

Table 22

Dayton and Macready Analysis: Likelihood and Misclassification Parameter Estimates for Hierarchies One to Eight (Unrestricted Sample)

Hierarchy	Guessing Parameter Estimate	Forgetting Parameter Estimate	Maximum Likelihood Estimate	Degrees of Freedom	Significance of Data
1	.50	1.00	336	2	0.000
2	.00	.06	47.6	9	0.000
3	.22	.00	51.9	23	0.001
4	.28	.04	15.2	8	0.055
5	.20	.00	38.1	22	0.018
6	.20	.00	38.3	21	0.012
7	.18	.00	31.6	21	0.064
8	.19	.00	31.8	20	0.046

are both consistent with the data, the hierarchy containing more skills should be chosen because it contains more information. In this case, therefore, Hierarchy Seven appears to be the more appropriate. The superiority of Hierarchy Seven over the other five-skill alternatives may be further tested by examining the significance of the difference between its likelihood ratio from the Dayton and Macready test, relative to each of the alternatives. The results of applying this test are given in Table 23. Δ represents the difference between the likelihood functions. The significance of the difference is determined by reference to a chi square table. The number of degrees of freedom is equal to the difference between the degrees of freedom for the two hierarchies when they were considered separately.

The results shown in Table 23 confirm that Hierarchy Seven is more appropriate than Hierarchies Three, Five and Six. Very little difference is observed, however, between it and Hierarchy Eight.

Repetition of the above analysis with the data restricted by removing those subjects who could not be clearly classified as masters or non-masters of the skills involved yielded the results represented in Table 24. When the data were restricted in this way all eight hierarchies were consistent with the data. The most consistent hierarchy was Hierarchy One, which formed a perfect fit. Hierarchy Four was again superior to Hierarchy Two, although the difference in this instance was significant only at the ten percent level. In all three of these hierarchies the values of the misclassification parameters were negligible. In the remaining five hierarchies, all containing five skills, the guessing parameter was large (.20 to .28) although the forgetting parameter was again minimal. Again Hierarchy Seven

Table 23

Likelihood Ratio Difference for Hierarchy Seven Against Hierarchies Three, Five, Six and Eight as Alternatives

Alternative Hierarchy	Δ	df	Significance of Data
3	20.3	2	.000
5	6.5	1	.011
6	6.7	0	
8	0.2	1	.655

Table 24

Dayton and Macready Analysis: Likelihood and Misclassification Parameter Estimates for Hierarchies One to Eight (Restricted Sample)

Hierarchy	Guessing Parameter Estimate	Forgetting Parameter Estimate	Maximum Likelihood Estimate	Degrees of Freedom	Significance of Data
1	.00	.00	0.00	2	1.000
2	.00	.01	10.6	9	0.304
3	.28	.00	18.7	23	0.719
4	.00	.01	7.9	8	0.443
5	.23	.00	16.05	22	0.813
6	.23	.00	16.10	21	0.764
7	.20	.00	13.63	21	0.885
8	.20	.00	13.63	20	0.849

provided the best fit. However the difference between the value of this statistic for Hierarchy Seven relative to each of the others was never significant at the five percent level with the possible exception of that for Hierarchy Six. Hierarchy Three was weaker than Hierarchy Seven at about the eight percent level. Moreover the guessing factor was worst for Hierarchy Three. Hierarchy Eight was again very similar to Hierarchy Seven in its goodness of fit to the data.

In summary it is suggested that the most appropriate hierarchy of the eight tested is Hierarchy Seven, because it offers a good fit to the data and is the best of larger hierarchies to do so. Nevertheless, the large value of guessing parameter is a concern. A smaller hierarchy, Hierarchy Four, also offers a good fit, and has negligible values for its misclassification parameters. However, the additional information given in Hierarchy Seven suggests the latter to be more appropriate. Whether it is in fact superior to Hierarchy Eight is open to some question. It may be best to compromise here and to include the additional connection in Hierarchy Seven over Hierarchy Eight as tentative. This may be represented in Hierarchy Nine:

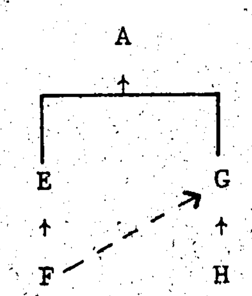


Figure 14. Hierarchy Nine as a Compromise Between Hierarchies Seven and Eight.

Several interesting implications arise from consideration of the results of applying the Dayton and Macready method to the data. Firstly, the superordinate relationship of Skill E (mass: equal # moles) over Skill H (molar mass as a ratio) was predicted in the originally hypothesized hierarchy and supported by the results of application of the White and Clark test and the ordering-theoretic method. However, the better fit of Hierarchies Seven and Eight (which do not include a connection between Skills E and H) relative to the fit of Hierarchies Five and Six (which do include this connection) to the same data is evidence against the existence of the necessity of hierarchical connection between Skills E and H.

A similar situation is found in the relationships between Skills G (mole: molar mass, L particles) and E (mass: equal # moles) and Skills G and F (mass: ≠# moles), respectively. According to the results of applying the White and Clark test and the ordering-theoretic method to the data, Skill G is superordinate both to Skill E and Skill F, a relationship which is in each case contrary to the expectations expressed in the hypothesized hierarchy. Application of the Dayton and Macready method to the same data allows a direct test of the relationship between Skills G and E within the context of the hierarchy as a whole, by comparison of the goodness of fit of Hierarchies Seven and Three to the same data. The much greater fit of Hierarchy Seven clearly indicates that Skill E is not a necessary prerequisite of Skill G. The marginal nature of the connection between Skill G and Skill F has already been discussed. Although Skill G as superordinate to Skill F represents an inverse relationship to what was hypothesized originally, the relationship is not entirely illogical in either direction. Skill G requires an understanding of the term mole in terms of molar mass and also in terms of the number of particles represented. The items testing this skill require recall of the Avogadro Number (L), the ability to identify that a given mass of a substance represents the molar mass (and implicitly to calculate the molar mass), and the ability to correctly apply these two parameters and to distinguish between their use if necessary in a given problem. Skill F requires the inter-conversion of mass of a substance and number of moles represented by that mass. In deriving the hypothesized hierarchy the position was taken that the learner should know what is meant by the term mole if he is to be able to calculate molar masses. In retrospect it is clear that while it may be desirable for meaningful learning that the learner should have an understanding of the term mole before he learns how to calculate molar masses, it is not essential. It is quite possible that the learner may be able to calculate molar masses algorithmically whether or not he understands the significance of the term molar mass. Further, it is equally clear in retrospect that the learner must be able to calculate molar masses if he is to demonstrate his understanding of the term mole as it relates to the number of particles in a given mass of a substance. What is illustrated here perhaps is the difference between the need to have grasped some idea of the meaning of a concept and the ability to apply that knowledge. The latter, which is illustrated by the items testing Skill G, requires the demonstration of an intellectual skill. Skill G is therefore seen as the skill needed to demonstrate understanding of the term mole. Skill F represents one of several components necessary to the correct application of Skill G. Viewed from this position it is not surprising that the data are at least quite suggestive that the ability to calculate molar masses is a prerequisite to the ability to demonstrate understanding of the term mole.

A further implication of the results of applying the Dayton and Macready method to the data is that the hierarchies which fit the data best (Hierarchies Four, Seven, and perhaps Nine) do not coincide exactly with those (Hierarchies One, Two and Three) derived by application of the White and Clark test and the ordering-theoretic method. It is apparent that consideration of skills in pairs does not necessarily lead to identification of the same hierarchy as when the skills are considered collectively. It is suggested that consideration of intact hierarchies is more appropriate because there is a greater chance of anomalous connections when skills are considered in pairs and the resulting connections combined to form a composite hierarchy. It is possible that the primary function of the White and Clark test and the ordering-theoretic method may become to sift data so that possible hierarchies may be suggested which may then be further tested by the Dayton and Macready method. However, the suggestion made in Chapter 2 that the Dayton and Macready method requires further refinement is reiterated at this time.

Before making a final interpretation of the results of the foregoing analysis one further important amendment may be made to Hierarchy Nine. From Table 15 (p. 162) it is evident that Skill C (mole: relative # particles) appears to be superordinate to Skill A (mass: relative # particles) when the White and Clark test is applied to the data, even at the most stringent level when only exceptions due to errors of measurement are allowed. It is not possible to justify this result in terms of the logic of the connection, even retrospectively. Further, Skill D (mole: # particles) appears to be empirically equivalent to Skill A (mass: relative # particles). Again, this result is surprising in view of the logical relationships postulated to exist between these skills when the hypothesized hierarchy was derived. One possible reason for these anomalous findings could be the extensive use of algorithmic procedures, the application of which would tend to negate an attempt to validate a logically derived hierarchy. However, analysis of the subjects' test papers indicated that the use of algorithmic procedures was limited to subjects in one class in one school, representing slightly less than ten percent of the sample. Moreover, this group of subjects accounted for only a very minor proportion of the anomalous responses.

Further examination of the subjects' test papers suggested the possibility that some subjects arrived at correct answers to test items for Skill A (mass: relative # particles) by applying Skill D (mole: # particles) without applying Skill C (mole: relative # particles), while others applied Skill C and only implicitly Skill D which was empirically subordinate to it. This may be illustrated by considering Item 19 testing Skill A in the Final Test. The subject is required to calculate the mass of ammonia (NH_3) containing twice as many molecules as 32 grams of oxygen (O_2). He may reason as follows:

32 grams of O_2 = 1 mole of O_2.

1 mole of O_2 = 6.02×10^{23} molecules of O_2.

6.02×10^{23} molecules of NH_3 = 1 mole of NH_3.

$2 \times 6.02 \times 10^{23}$ molecules of NH_3 = 2 moles of NH_3.

2 moles of NH_3 = 34 grams of NH_3.

This route involves application of Skill D (mole: # particles) but not Skill C (mole: relative # particles). Alternatively, he may reason:

32 grams of O_2 = 1 mole of O_2.

The number of molecules of a compound is proportional to the number of moles present.

Twice as many molecules requires twice as many moles.

Two moles of NH_3 are required to contain twice as many molecules as one mole of O_2.

Two moles of NH_3 = 34 grams.

This route involves Skill C but not (directly) Skill D.

Each of these routes was used by subjects in the present study. As a consequence Skills C and D might appear to be more difficult relative to Skill A than is really the case. This suggests the possibility that Skills C and D might yet be appropriately placed in an hierarchy leading to Skill A as the superordinate skill. The relationship of Skills C and D to those skills already shown to be subordinate to Skill A was therefore considered first. Examination of Table 15 indicates that, with the exception of Skill E at the 00 level, all of the skills which are subordinate to Skill A are also subordinate to Skills C and D. Further, Skill D is subordinate to Skill C. These considerations lead to the possibility of Hierarchy Ten (following page).

It is not possible to use the White and Clark test to test the reasonableness of the disjunctive relationship which Hierarchy Ten suggests for the connections between Skills A, C and D. The Dayton and Macready method does allow for such a test. This was applied

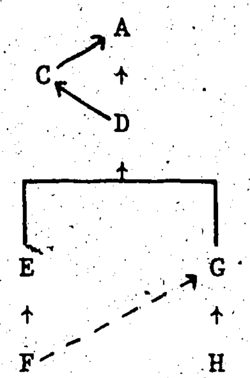


Figure 15. Hierarchy Ten.

without eliminating those subjects with central scores. The guessing parameter was found to be 0.18, the forgetting parameter was 0.02, and the value of the maximum likelihood function was 55.67 with 52 degrees of freedom. This value indicates that Hierarchy Ten is not rejected at the .05 level of significance. Because it contains more useful skills than do Hierarchies One to Nine and also offers a good fit to the data, Hierarchy Ten is considered to be the most appropriate hierarchy. It is therefore considered to be the psychometrically valid hierarchy and is accordingly re-stated in more detail in Figure 16 in terms of its component skills.

Transfer of Learning Within the Psychometrically Validated Hierarchy

Research question five is concerned with the existence of transfer of learning from subordinate to related superordinate skills in the psychometrically validated hierarchy. Hypothesis two was derived from this question. Hypothesis two states that '*no significant relationship exists between acquisition of subordinate skills and acquisition of related superordinate skills in the validated hierarchy.*'

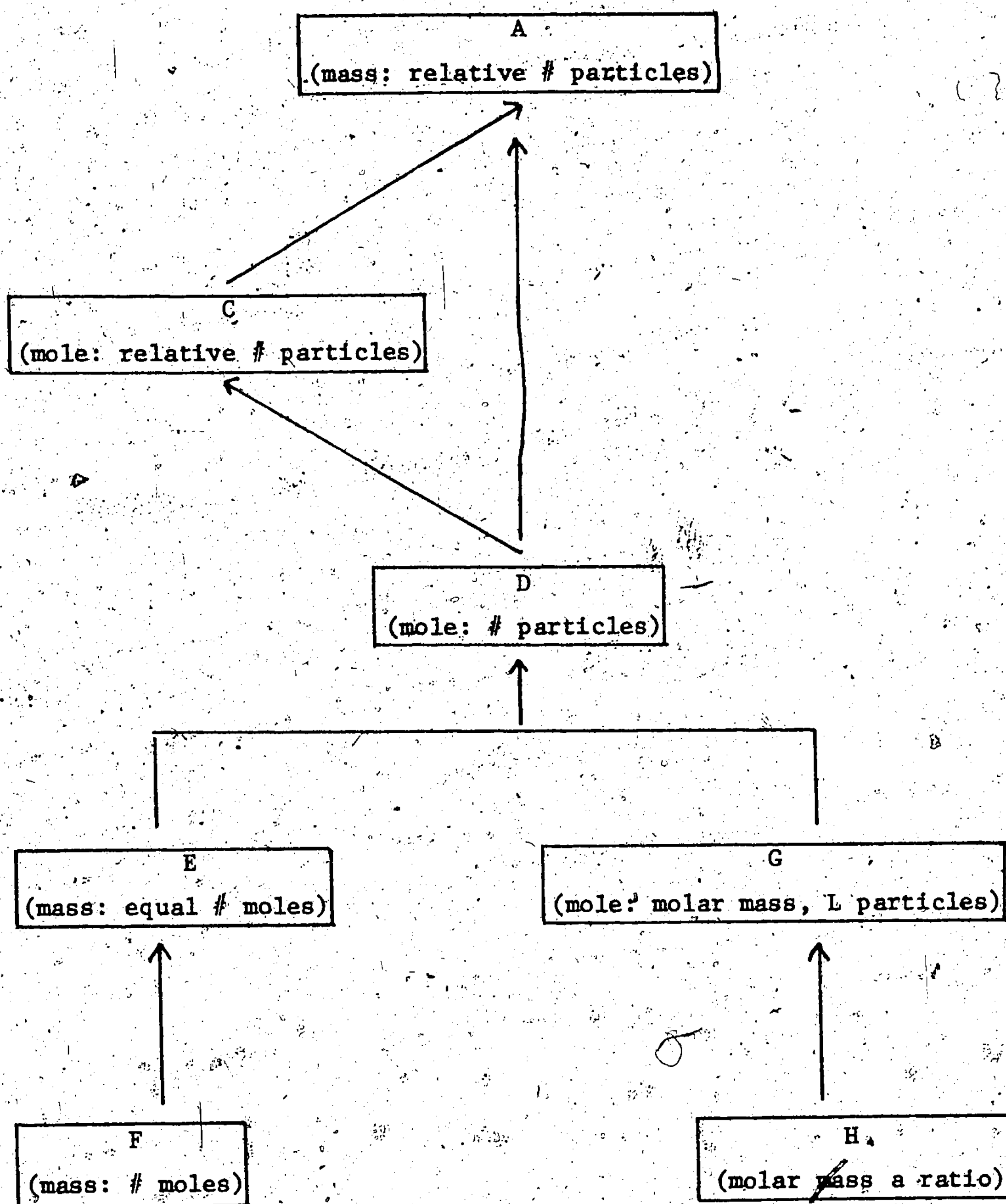


Figure 16. The psychometrically validated hierarchy.

As indicated in Chapter 2, Gagné's index, proportion positive transfer, and others like it is not considered to be an acceptable test of the degree of transfer between skills in an hierarchy. A more direct method is to remediate for missed subordinate skills and then compare achievement of the superordinate skill for the remediated group versus a control group not receiving remediation. This procedure was initially adopted in the present study. Remedial units relating to the development of the subordinate skills in the hypothesized hierarchy were administered within the instructional period. The remedial sub-sample was comprised of 107 subjects in four intact classes which were randomly assigned before commencement of instruction to receive the remedial units. Details relating to the administration of the remedial units were described fully in Chapter 5 (p.128). Briefly, Remedial Unit One representing Skills G and H was administered immediately after Quiz One which tested Skills G and H, and Remedial Unit Two representing Skills C, D, E and F was administered immediately after Quiz Two which tested Skills C, D, E and F. Both remedial units were administered as take-home assignments which were returned with completed exercises at the beginning of the next instructional period in chemistry. This procedure resulted in administration of remedial units to all subjects in the remedial group including those who did not need remediation. This procedure was designed to minimize interference with normal teacher and student activities. However, those in each group who successfully exhibited a particular remediated skill in the prior quiz were deleted from the sample in any analysis relating to remediation of that particular skill. For the purpose of such analysis those subjects needing and receiving remediation for the particular skill concerned were designated 'remedial' and those subjects needing but not receiving remediation were designated 'non-remedial.'

Before determining the relationship between gain of remediated skills and acquisition of the superordinate skill, it was necessary to ascertain if the remedial group exhibited greater gain of each of the remediated skills, relative to the control group not offered remediation. To establish this the quizzes were considered as pre-tests and the Final Test as post-test. The numbers of subjects in the remedial and non-remedial groups gaining and failing to gain the remediated skills are reported in Table 25. A chi square test was applied to these data to determine the statistical significance the difference in proportionate gains for these two groups. The results are presented in Table 26. Skill H was not included in the analysis because only three percent of the sample failed to exhibit it in Quiz One.

In all cases the remedial group outgained the non-remedial group. However, ..e results in Table 26 illustrate that in no case was the differen :e significant. Hence the intended test of transfer could not be applied. The problem lay not with failure of the remedial group to show substantial gain but with the fact that the non-remedial group also gained substantially, thereby reducing the relative gains made by the remedial group. This result is at first surprising, but is consistent with Wiegand's (1969) study in which testing of subordinate skills promoted gain of superordinate skills. Nevertheless, in designing the present study it had been hoped that

Table 25

Numbers of Remedial and Non-Remedial Subjects Gaining and Failing to Gain Remediated Skills Between Quiz and Final

Skill	Remedial		Non-Remedial	
	Failed to Gain	Gained	Failed to Gain	Gained
G	12	32	27	42
F	6	16	6	11
E	13	28	23	24
D	46	17	79	19
C	47	14	74	11

Table 26

Results of a Chi Square Test to Determine the Effectiveness of Remediation of Skills C, D, E, F and G

Remediated Skill	χ^2	df	Significance
G	2.82	1	>.05
F	0.29	1	>.05
E	2.69	1	>.05
D	2.51	1	>.05
C	1.27	1	>.05

the effect of remediation would have been sufficiently greater than the effect of testing for the former to significantly outweigh the latter. Unfortunately, for the transfer design in the present study this was not found to be so.

In order to provide some evidence of the existence of positive transfer within Hierarchy Ten an alternative procedure was applied which, although not as strong as that which was intended, is considered to provide useful information. The following steps were involved in this alternative analysis:

1. Those skills in Hierarchy Ten which are directly subordinate to any other skills in the hierarchy were identified.
2. Subjects who failed to exhibit any of these subordinate skills in the relevant quiz were identified for each particular skill and for groups of these skills where a group was directly subordinate to any particular skill(s) in Hierarchy Ten. The size of the sub-sample selected in this way varies from skill to skill. No distinction was made between subjects previously designated remedial or non-remedial as the question to be answered in the alternative procedure related to the relationship of gain of subordinate skills between quiz and Final Test and exhibition of related superordinate skills in the Final Test.
3. The performance in the Final Test of each of the subjects identified in step 2 was determined with respect to whether or not the subject exhibited the subordinate skill(s) he failed to exhibit in the quizzes. These subjects were labelled 'Gain' if they exhibited the skill(s) in the Final

Test and 'No-Gain' if they failed to exhibit the skill(s) in the Final Test.

4. The performance on the Final Test of each of the subjects identified in step 2 was determined with respect to any skill superordinate to any of the subordinate skills he failed to exhibit in the quizzes. Based upon performance on this superordinate skill subjects were designated as 'Pass' or 'Fail' for the particular skill.
5. The significance of the relationship between Gain/No-Gain and Pass/Fail was determined by application of a chi square test with one degree of freedom in each case.

Two exceptions were made to the above procedure. Once again the connection between Skills G and H was not considered because the proportion of subjects failing to exhibit Skill H was too small to allow meaningful interpretation of the results. Secondly, because of their disjunctive relationship to Skill A (mass: relative # particles) Skill C (mole: relative # particles) and Skill D (mole: # particles) were treated as a pair in which gain represented a change of status from failure for both in Quiz Two to success for either in the Final Test. This is not totally satisfactory because some subjects who show gain of Skill D but not Skill C may try, almost certainly unsuccessfully, to apply Skill C to the solution of items testing Skill A. The net result is to depress the observed transfer. When Skills E (mass: equal # moles) and G (mole: molar mass, L particles) were considered as conjointly subordinate to Skill D (mole: # particles), failure was considered as failure to exhibit either Skill E or Skill G and success was considered as success on both skills.

The results of applying the alternative test of transfer in the manner described above are presented in Table 27. These results indicate that subjects who gained subordinate skills between the quizzes and the Final Test exhibited significantly greater achievement of related superordinate skills in the Final Test. This finding is consistent with the existence of positive transfer of learning between related skills in Hierarchy Ten. It is therefore supportive of the hierarchy.

The above finding might imply that remediation of subordinate skills until mastery would be beneficial before learners are required to learn related superordinate skills. However, the finding presented earlier in the present chapter that the remedial units used in the present study did not significantly enhance the learning of the subordinate skills does not support the above implication. It is possible that the remedial units themselves were ineffective. However, examination of the proportion of subjects in the remedial group who needed remediation and actually gained the skill(s) involved tends to deny this explanation.

Two other explanations are also consistent with the observed gains. The first, which has been discussed already, is that testing of skills within the instructional period is sufficient to induce gain of skills which subjects failed to master at the time of initial instruction of those skills. However the fact that the remedial subgroup exhibited greater gain of <u>each</u> subordinate skill, even though the difference was not statistically significant, may be considered to ameliorate this 'effect of testing' explanation to some extent.

A second alternative explanation for the transfer data is the

Table 27

Alternative Test of Transfer from Subordinate to Superordinate Skills

Connection	No-Gain Fail (N)	No-Gain Pass (N)	Gain Fail (N)	Gain Pass (N)	χ^2	df	Significance
C, D to A	73	37	9	15	5.75	1	<0.05
G to A	72	34	31	38	8.20	1	<0.01
D to C	119	5	21	12	24.97	1	<0.001
E to D	39	5	33	17	5.49	1	<0.05
G to D	89	8	28	41	48.31	1	<0.001
E, G to D	101	17	24	42	44.86	1	<0.001
F to E*	12	2	12	15	4.88	1	<0.05
F to G*	14	0	15	12	6.78	1	<0.05

*Yates correction applied

possibility of downward transfer (Cotton et al., 1978) in which the learning of later skills enhances the learning of partially learned earlier skills. Provision for a sub-group of subjects in the present study who were not exposed to remediation nor testing until the end of the instructional period would have allowed testing of this explanation. However, no provision was made for this.

In summary, the transfer data supports Hierarchy Ten although no significant support is given to the suggestion that remediation of subordinate skills enhance learning of related superordinate skills. It is appropriate at this time to discuss the structure of Hierarchy Ten, which is considered to be the hierarchy validated by the present study. This discussion follows immediately.

The Structure of the Validated Hierarchy

Hierarchy Ten (p. 191) exhibits similarities to and differences from the hypothesized hierarchy (p. 116). Of the seven skills hypothesized to be subordinate to Skill A (mass: relative # particles) only Skill B (mass: # particles) was not found to be subordinate. The reason for this exception is not clear, especially as a parallel relationship between Skill C (moles: relative # particles) and Skill D (moles: # particles) was substantiated empirically.

A number of other relationships hypothesized to exist between pairs of skills were substantiated. The ability to exhibit Skill C (moles: relative # particles) and Skill D (moles: # particles) required, as hypothesized, the ability to demonstrate understanding of mole in terms of molar mass and the number of particles represented therein (Skill G) and also an understanding of molar mass as a ratio

(Skill H). In turn Skill G was superordinate to Skill H, as hypothesized. Finally, Skill E (mass: equal # moles) was superordinate to Skill F (mass: # moles).

However, as well as the unexpected lack of relationship between Skill A and Skill B, several other surprising results were obtained. It has been anticipated that Skills C and D (relating numbers of moles to numbers of particles present) would develop independently of Skills E and F (relating numbers of moles and masses). However, Skills C and D were found to be clearly superordinate to Skills E and F. A possible reason for this is that it might be easier for the learner to relate moles to masses than to numbers of particles because masses may be more easily related to direct experience. Expressed another way, the relationship between these pairs of skills may be psychological rather than logical.

A further unexpected relationship between Skills F and G in which the expected order is inverted with Skill G superordinate to Skill F has already been discussed (pp.186, 187). To reiterate briefly, the learner does not necessarily have to be able to exhibit an understanding of the term mole (Skill G) before he can interconvert the mass of a substance and the number of moles represented by this mass (Skill F). Indeed, it appears more likely that the learner must be able to calculate molar mass before he learns the meaning of the term mole.

Finally, Skill H (molar mass as a ratio) is not subordinate to Skills E and F. Again, in retrospect, this is not difficult to rationalize. What this result suggests is that the learner can apply the numerical values of molar masses in some situations without necessarily understanding what is implied by these values. However, it is also noted that Skill H is subordinate to Skill G and is therefore also subordinate to Skills A, C and D. Hence it may be necessary to expose the learner to the meaning of molar mass as a ratio early in the treatment of the mole. However, this suggestion is tentative because it was not possible to test for transfer from Skill H to related superordinate skills because too few subjects failed Skill H originally.

Before turning to the Research questions six and seven which relate to learner developmental level, one final question will be considered relating to the existence of an hierarchy of skills leading to the mole concept. This is the question of when to test, and in the present study whether the Final Test data are more appropriate than the quiz data.

A Comparison of Hierarchical Connections Validated Within and After the Instructional Period

White (1973) has criticized those studies in which the existence of hierarchical connections has been tested after instruction, rather than during instruction. He argues that the former approach confounds the results because the effect of differential forgetting of skills is not controlled. The argument has already been made in the present report that it is possible that some subjects may acquire skills as a result of further instruction not directly designed to promote those particular skills. Depending on the pattern of such gains a variety of spurious outcomes are possible. For example, suppose that in the present study a sample achieves a significant gain on a particular skill (say, Skill G) between testing on Quiz One and Quiz Two. Further, suppose that after Quiz Two there is no further gain of Skill G or

Skill D (which was tested in Quiz Two). This would probably result in the identification of more exceptions to an hierarchical connection in which Skill D is tested as superordinate to Skill G if the quiz data was used for the test than if the Final Test was used. If the additional gain of Skill G was made after Quiz Two and again no further gain was made for Skill D there would probably be little difference in the number of exceptions to the hierarchical connection if the test was made on the quiz data or those from the Final Test. The problem is that it is not possible to tell at what time any gains were actually made.

To test if there is any difference in the present study when the tests for hierarchical connections are made using quiz results rather than those from the Final Test the data from those subjects who did not receive remediation were used. In order to test if this non-remedial sub-group was representative of the complete sample the ordering-theoretic method was first applied at the 5% level to the Final Test data for this sub-group, and the hierarchical connections identified by this analysis were then compared to those identified for the complete sample for the Final Test data. In each case the data were restricted by removing central scores in the manner described for the earlier application of the ordering-theoretic method. Table 28 indicates the percentage of exceptions for each possible connection in each direction when the data from the non-remedial sub-group on the Final Test were analyzed. The connections identified from this analysis are summarized in Table 29. These results are in substantial agreement with those obtained from a similar analysis for the main sample (Tables 17 and 19, pp. 170, 173). Only one difference was observed.

Table 28

Ordering-Theoretic Method: Percentage of Exceptions to Hierarchical Connections (Restricted Non-Remedial Sub-Sample, Final Test)

		'Upper' Skill							
		A	B	C	D	E	F	G	H
	A	--	7.0	2.4	6.5	30	50	19	45
	B	7.5	--	4.1	0.9	32	45	13	44
	C	12	15	--	10	41	64	24	55
'Lower' Skill	D	7.4	2.8	0.9	--	32	48	10.9	41
	E	-	0.9	0	1.0	--	9.2	4.2	9.0
	F	0	0	0	0	0	--	1.8	0.9
	G	6.4	4.3	1.1	1.1	9.4	19.5	--	21
	H	0.9	2.6	0	1.0	2.7	4.5	1.0	--

Table 29

Summary of Hierarchical Connections Derived by Applying the Ordering-Theoretic Method at the 5% Level (Restricted Non-Remedial Sub-Sample, Final Test)

A	B	C	D	E	F	G	H
E	D	A	B	F	H	E	F
F	E	B	E	H		F	
H	F	D	F			H	
	G	E	G				
	H	F	H				
		G					
		H					

Skill A was superordinate to Skill G using the data from the main sample while no hierarchical relationship between these skills was found in the data from the remedial sub-sample. However, even this one difference was only marginally accepted for the main sample and marginally rejected for the sub-sample. In contrast, as may be inferred from Tables 30 and 31 the same sub-sample exhibited a quite different pattern of connections on the quizzes compared to the Final Test. Eight differences were observed. Connections B (mass: # particles)/A (mass: relative # particles) and G (mole: molar mass, L particles)/A (mass: relative # particles) were observed on the quiz data but not on the data from the Final Test. Each of these connections is difficult to justify in terms of the logic of the subject matter. Six other connections, namely A (mass: relative # particles)/E (mass: equal # moles); B (mass: # particles)/G (mole: molar mass, L particles); B (mass: # particles)/D (mole: # particles); D (mole: # particles)/G (mole: molar mass, L particles); G (mole: molar mass, L particles)/E (mass: equal # moles) and H (molar mass as a ratio)/F (mass: # moles), respectively, were observed only for the Final Test data. It is noteworthy that the first four of these six connections which were exhibited in the Final Test data but not in the quiz data were hypothesized in the originally hypothesized hierarchy and that the fifth connection found (G/E), although not hypothesized, has already been rationalized in terms of the logic involved. Moreover, for each of these five connections the lower skill was tested on a prior quiz to that used to test the upper skill. This is consistent with the earlier suggestion that anomalies may occur because skills are gained between the time they are tested and the time that later skills are tested.

Table 30

Ordering-Theoretic Method: Percentage of Exceptions to Hierarchical Connections (Restricted Non-Remedial Sub-Sample, Quizzes)

		'Upper' Skill							
		A	B	C	D	E	F	G	H
	A	--	4.1	1.2	5.4	21	52	3.8	60
	B	10	--	2.3	4.3	26	68	7.9	67
	C	7.2	6.9	--	7.6	27	61	14	67
'Lower' Skill	D	8.7	6.5	1.0	--	20	53	10	62
	E	5.1	2.6	1.0	3.1	--	2.9	5.1	39
	F	1.1	1.1	0	0	0	--	0	11
	G	15	7.9	3.6	11	20	47	--	53
	H	1.1	2.1	0	0	1.1	3.3	1.0	--

Table 31

Summary of Hierarchical Connections Derived by Applying the Ordering-Theoretic Method at the 5% Level (Restricted Non-Remedial Sub-Sample for (a) the Final Test, (b) the Quizzes)

	A	B	C	D	E	F	G	H
Final Test	E F H	D E F G H	A B D E F G H	B E F G H	F H	H	E F H	F
Quizzes	F H	E F H	A B D E F G H	B E F H	F H	E H	A F H	

To summarize, White's contention that delaying testing until after instruction is complete results in different hierarchical connections is supported by the data in the present study. Whether he is correct in advocating testing only during the instructional period is challenged by the same data. For each skill the number of subjects showing gain of the skill between the initial testing and the Final Test exceeds the number showing loss of the same skill between these two occasions. Hence, forgetting does not appear to be a major factor. Spurious results appear more likely if data from the quizzes are used rather than the data from the Final Test.

Learner Developmental Level and Acquisition of the Skills in the Validated Hierarchy

Research question six asks whether a significant relationship exists between learner developmental level and acquisition of each skill in the validated learning hierarchy. This question is re-stated in the form of the null hypotheses 3(a) and 3(b), which state that no significant relationship exists between achievement of each skill in the validated hierarchy and scores on the Test of Developmental Level and the Skemp Test respectively. These hypotheses were tested by application of correlation and regression analyses.

Pearson product-moment correlations were first determined between total scores on the Test of Developmental Level (DL), the Skemp Test (SK) and each of the Final Test sub-tests for the seven skills represented in Hierarchy Ten. The matrix representing these correlations is presented in Table 32 (it is noted that a similar analysis for each individual classroom group did not indicate that any group was substantially different to any of the others in the pattern of correla-

Table 32

Pearson Product-Moment Correlations Between Final Test Scores for Skills A, C, D, E, F, G and H, the Test of Developmental Level (DL), the Skemp Test (SK)
N = 209

	DL	SK	Skill A	Skill C	Skill D	Skill E	Skill F	Skill G	Skill H
DL	1.00								
SK	.48	1.00							
Skill A	.30	.35	1.00						
Skill C	.27	.35	.67	1.00					
Skill D	.36	.41	.65	.74	1.00				
Skill E	.30	.37	.61	.42	.44	1.00			
Skill F	.29	.29	.45	.37	.36	.51	1.00		
Skill G	.30	.31	.56	.61	.61	.37	.35	1.00	
Skill H	.09**	.15*	.23	.24	.24	.21	.21	.24	1.00

*Significant at .05 level.

**Not significant

Note: All other correlations significant at 0.001 level.

tions obtained). In all cases the correlations obtained for the Skemp Test with the chemistry skills tests were greater than or equal to those obtained for the Test of Developmental Level with the same skills. This finding appears to suggest that the ability to exhibit second-order thinking is more important to the acquisition of the intellectual skills represented in the development of the mole concept than are the more generalized intellectual competencies represented in the Test of Developmental Level. However, application of a 't' test (Ferguson, 1976, p. 185) to determine the significance of the difference between these correlations indicated that in no case was the difference significant at the 0.05 level.

Scores on both the Test of Developmental Level and the Skemp Test exhibited correlations with the scores on the tests for six of the skills (Skills A, C, D, E, F and G) in Hierarchy Ten which were significantly different from zero at the 0.001 level of confidence. With respect to these six skills Hypotheses 3(a) and 3(b) are therefore rejected. The remaining skill, Skill H, did not correlate significantly with scores on the Test of Developmental Level and exhibited a correlation of 0.15 (significant at 0.05 level) with scores on the Skemp Test. Hypothesis 3(a) is therefore rejected and Hypothesis 3(b) tentatively accepted for Skill H. In general, research question six is answered affirmatively for the skills represented in Hierarchy Ten. A significant relationship exists between the developmental level of the learner and acquisition of each skill in the validated hierarchy. However, the magnitude of this relationship can be considered no more than moderate in each case, with a range between 0.27 and 0.41.

Despite the relatively low correlations between scores on both the Test of Developmental Level and the Skemp Test with the tests for the Chemistry skills represented in Hierarchy Ten it is possible that collectively the two developmental tests may contribute more substantially to the scores on the tests for the chemistry skills. The moderate correlation (r = 0.48) between the Test of Development Level and the Skemp Test supports this possibility in that it is consistent with the interpretation that the two tests may be measuring some different aspects of formal operational thinking. To assess their combined contribution to the variance of the scores for the tests for each of the skills in Hierarchy Ten a series of stepwise regression analyses were performed with the test score for each particular skill in turn as dependent variable and the Skemp Test and Test of Developmental Level scores as predictors in each case. In all cases the Skemp Test scores entered first, as would be expected from the relative values reported in Table 32 for the correlation between each of these tests and the chemistry skills scores. The results of the regression analyses are reported in Table 33. They indicate that the proportion of the variance in scores on the tests for the skills represented in Hierarchy Ten which can be accounted for by the Skemp Test and the Test of Developmental Level collectively is relatively small in all cases, and in no instance exceeds 18 percent of the variance. This proportion of the variance is sufficiently large to encourage the suggestion that learner developmental level has some bearing on acquisition of the skills leading to development of the mole concept, but is not large enough to suggest that developmental level is of major importance to acquisition of these skills.

Correlations between scores on tests of skills exhibiting a direct

Table 33

Proportion of the Variance in the Chemistry Skills Tests Accounted for by (1) the Skemp Test, and (2) the Skemp Test (SK) and the Test of Developmental Level (DL) Combined

Skill	R^2 (SK)	R^2 (SK + DL)
A	0.124	0.149
C	0.122	0.137
D	0.166	0.178
E	0.139	0.159
F	0.098	0.129
G	0.097	0.125
H	0.021	0.021

hierarchical relationship to one another in Hierarchy Ten were all higher than the correlations observed between each superordinate skill and the scores on the developmental level tests. The relative importance of availability of subordinate skills and learner developmental level is the focus of the final research question addressed in this study. This question is pursued in the section which follows.

The Relative Importance of Developmental Level and Availability of Subordinate Skills to the Acquisition of Superordinate Skills

Research question seven is concerned with the relative importance of developmental level and availability of subordinate skills to the acquisition of related superordinate skills. Hypothesis 4 stated in the null form suggests that there is no difference between these two parameters as predictors of achievement of superordinate skills.

The inter-correlations between the test scores for the skills represented in Hierarchy Ten were reported in Table 32. The values of these correlations are generally much higher than those between the same chemistry skills tests and the Skemp Test and Test of Developmental Level, respectively. To test whether the correlations between test scores for hierarchically related skills are significantly greater than the correlations between the superordinate skill scores and scores on each of the developmental tests a 't' test was applied (Ferguson, 1976, p. 185). The results of applying this test are indicated in Table 34.

The correlations reported in Table 34 indicate that the learner's ability to compare masses of substance in terms of the relative numbers of particles represented (Skill A) is related more strongly to his

Table 34

Significance of the Differences Between Correlations of Test Scores for Hierarchy Skills with Scores on Tests for Immediately Subordinate Skills Compared to Their Correlations with Scores on (1) the Test of Developmental Level (DL) and (2) the Skemp Test (SK)
N = 209

Superordinate Skill	Subordinate Skill (SUB)	Correlation With SUB	DL	SK	t	P
A	C	.67	.30		6.05	<.001
A	C	.67		.35	5.55	<.001
A	D	.65	.30		6.65	<.001
A	D	.65		.35	5.29	<.001
C	D	.74	.27		9.04	<.001
C	D	.74		.35	7.74	<.001
D	E	.44	.36		1.06	N.S.
D	E	.44		.41	0.43	N.S.
D	G	.61	.36		3.97	<.001
D	G	.61		.41	3.25	<.001
E	F	.51	.30		3.01	<.01
E	F	.51		.37	2.05	<.05
G	H	.24	.30		0.70	N.S.
G	H	.24		.31	0.83	N.S.
G	F	.35	.30		0.66	N.S.
G	F	.35		.31	0.53	N.S.

ability to convert moles to relative numbers of particles (Skill C) or actual numbers of particles (Skill D) than to his level of intellectual development. In turn, acquisition of Skill C is more strongly related to acquisition of its immediate subordinate Skill D than to developmental level. Similarly, acquisition of Skill D is more strongly related to one of its immediate subordinates Skill G (mole: molar mass, L particles) than to developmental level, although the correlation of Skill D with its other immediate subordinate Skill E (mass: equal # moles) is not significantly greater than the correlations of Skill D either with the Test of Developmental Level or the Skemp Test. Skill E follows the general pattern described above by correlating significantly better with its immediate subordinate Skill F (mass: # moles) than with either developmental level test. Only Skill G (mole: molar mass, L particles) failed to correlate better with its immediate subordinates Skills H (molar mass as a ratio) and F (mass: # moles), respectively, than with the scores on the Test of Developmental Level and the Skemp Test. It is possible that the low failure rate for Skills H and F (approximately four percent and ten percent, respectively), decreased the variance of the scores for these skills sufficiently to significantly depress the correlations of these scores with those for Skill G, thereby making the correlations between scores for Skill G and each of the developmental level tests spuriously important.

In general the results discussed above suggest that the availability of specific intellectual skills is more important than the developmental level of the learner to his learning of the mc con- cept, a finding which is consistent with the Gagnéan hierarchical model of learning. To test this further the combined contribution of the Test of Developmental Level and the Skemp Test was compared to the contribution of immediately subordinate intellectual skills as predictors in regression equations in which the test scores for each skill superordinate to any other(s) in Hierarchy Ten were treated successively as criterion variable. The results of this analysis are presented in Table 35. These results further support the suggestion that the availability of related subordinate skills is more important than learner developmental level to achievement of the intellectual skills examined in the present study. This interpretation is reinforced when the contributions from the two developmental tests are added after the contributions from immediately subordinate skills have been entered into stepwise regression equations in which the same superordinate skills represent the criterion variables. The results of this analysis which are presented in Table 36 indicate that little is gained, especially for the three most superordinate skills (Skills A, C and D), by including the developmental contribution in the regression equations. Hypothesis seven, that no significant difference exists between the effectiveness of the developmental tests and subordinate skills as predictors of achievement of related superordinate skills is therefore rejected. The most important factor contributing to achievement of the mole concept is the availability of specific subordinate intellectual skills.

Summary of Results

In this chapter the research questions and hypotheses which form the crux of this study have been investigated. Briefly, the following major outcomes were identified: The hypothesized hierarchy was not

Table 35

Contributions (R^2) of (1) Scores on Tests of Immediate Subordinate Skills (SUB), (2) Scores on the Skemp Test (SK) and the Test of Developmental Level (DL) to the Variance of Test Scores for Superordinate Skills when (1) and (2) are Entered Separately into the Regression Equations
N = 209

Superordinate Skill	Subordinate Skills (SUB)	R^2 (DL + SK)	R^2 (SUB)
A	C, D	0.15	0.51
C	D	0.14	0.55
D	E, G	0.18	0.43
E	F	0.16	0.26
G	F, H	0.13	0.15

Table 36

Additional Contribution of Developmental Test Scores (DL + SK) to the Variance on Tests of Superordinate Skills after the Contribution from Scores on Tests of Immediate Subordinates has been Entered into the Regression Equation

N = 209

Superordinate Skill	Subordinate Skills	R^2 (SUB)	ΔR^2 (R^2 (total) - R^2 (SUB))
A	C, D	0.510	0.011
C	D	0.547	0.005
D	E, G	0.429	0.028
E	F	0.265	0.054
G	F, H	0.151	0.056

substantiated by any of the tests applied. Alternative hierarchies were suggested following the application of the White and Clark test, the extensiveness and structure of these hierarchies varying according to the degree of stringency applied to the test. Good agreement was found between these results and those obtained by applying the ordering-theoretic method to the same data. Application of the Dayton and Macready method to these data yielded a more extensive hierarchy containing all except one of the originally hypothesized skills. Because this larger hierarchy contained relationships between more skills and because it was derived by consideration of the whole hierarchy rather than by consideration of pairs of skills it was considered to be the most appropriate of those tested. It was confirmed by the existence of positive transfer of learning between related skills. Finally, learner developmental level was found to correlate only weakly with the skills in the hierarchy. The most important contributor to acquisition of the skills involved in the present study was the availability of related subordinate skills.

Chapter 7

SUMMARY, IMPLICATIONS AND RECOMMENDATIONS

Summary

The major purpose of the present study as initially conceived was to attempt to identify a valid learning hierarchy leading to an understanding of particular aspects of the 'mole' concept, a concept which is both of central importance in many high school chemistry courses and a source of difficulty to many high school students. Unfortunately, many of the methods which have been used by researchers to validate learning hierarchies have been the subject of cogent criticisms. Several recent methods have been developed in response to these criticisms, but these methods have not themselves been extensively analyzed in terms of their theoretical and empirical characteristics. Consideration of these characteristics for three methods and a comparison of the hierarchies 'validated' by them when applied to the same data became a primary focus of the study. This offers not only a potentially important contribution to the research literature concerning methods of hierarchy validation, but also enables greater confidence to be placed in the hierarchy validated in the present study. The final purpose of the study was to determine the relationship between learner developmental level and acquisition of the intellectual skills representing development of the mole concept, and further to evaluate the relative importance of developmental level and availability of subordinate skills to the acquisition of related superordinate skills.

A learning hierarchy for the mole concept was proposed by the investigator and modified after field testing and discussion with teachers. The resultant hierarchy represented the hypothesized hie rchy for the study. Four chemistry tests, each compounded from several smaller criterion-referenced tests, were developed to test the skills represented in the hypothesized hierarchy. Each was pilot-tested and modified as a result of the feedback gained. Three of the tests, labelled Quiz One to Quiz Three, were each administered by the co-operating teachers to their own classes in the next regular chemistry class following instruction of the skills represented in the particular test. Collectively these three tests tested all of the skills in the hypothesized hierarchy. These skills were again tested in the Final Test which was administered several days after completion of instruction relating to the skills in the hypothesized hierarchy. In addition to the chemistry tests two developmental level tests were administered by the investigator just prior to the commencement of instruction of the mole concept. The first of these, the Skemp Test, was considered to be a measure of the ability of the subject to perform second-order operations. The second consisted of a battery of three neo-Piagetian tasks representing different aspects of formal operational thought. In addition to the test instruments two written remedial units representing the skills tested in Quiz One and Quiz Two respectively were developed. These were administered immediately after these quizzes and returned in the next chemistry class period.

Three tests were applied to the Final Test data in order to determine the validity of the hypothesized hierarchy. These tests were the White and Clark test of inclusion, an 'ordering-theoretic' method developed by Bart and others, and a scaling method developed by Dayton and Macready. As a result of this procedure the hypothesized hierarchy was rejected but a number of other smaller hierarchies derived from the same set of skills were suggested as possible alternatives. Ultimately, after testing these hierarchies the arrangement of skills represented in Figure 15 (p. 191) was considered to represent a psychometrically valid hierarchy. This hierarchy contained seven of the eight skills represented in the hypothesized hierarchy. The validity of this hierarchy was confirmed by a test of transfer in which a significant relationship was found to exist between gain of subordinate skills and gain of related superordinate skills. The results obtained from application of the White and Clark test an the ordering-theoretic method were not identical but showed substantial agreement. The hierarchies derived from application of these methods were not entirely consistent with those validated by application of Dayton and Macready's procedures. In such cases the hierarchies validated by Dayton and Macready's procedure were accepted because they considered the validity of the hierarchy as a complete entity rather than by aggregating hierarchical connections between pairs of skills.

Correlations between scores on each of the developmental tests and the tests for each of the chemistry skills were generally significant but moderate. The largest correlation was only 0.43. Collectively the two developmental level tests accounted for no more than 18 percent of the variance in the scores for any of the tests for the chemistry skills. This was substantially less than the proportion of the variance in the same chemistry skills tests scores which was accounted for by scores on the tests of subordinate skills. When the developmental test scores were entered after scores for subordinate skills into regression equations in which the dependent variable in each case was a related superordinate skill the increase in explained variance was negligible for the three most superordinate skills and relatively small in all cases.

Implications

The study has implications for the methodology of learning hierarchy validation, the arrangement of instruction for the mole, and the relationship between Gagnéan and Piagetian theory.

Until generally acceptable methods are available for the validation of learning hierarchies consistent progress cannot be made in the arrangement of instruction of intellectual skills nor can the hierarchy model itself be properly articulated. Two of the methods applied in the present study, the White and Clark test and the Dayton and Macready method, offer much promise in this regard. The White and Clark test has significant advantages over earlier methods which consider skills in pairs. In particular it allows for errors of measurement and provides an appropriate test of significance. However, in its original form it has several limitations. It allows only for exceptions due to measurement error and is restricted in the number of test items which can be used. Further, it is restricted to consideration of skills in pairs. The first limitation has been removed by Linke's modification to allow for different degrees of stringency, although in turn it should be recognized that Linke's modification re-introduces an element of arbitrariness into the decision of hierarchical dependency between two skills. The second limitation has been removed by Griffiths and Cornish in the context of the present study. The modification proposed is considered to be an important outcome of the study. The third limitation is shared by most other methods which have been used in the validation of hierarchies, but is nevertheless an important limitation because the aggregation of pairs of skills which are hierarchically related may not lead to the same hierarchy as when that hierarchy is considered as a composite.

This latter limitation is removed in the method suggested by Dayton and Macready, which considers the pattern of responses in the composite hierarchy. Moreover, this method has an appropriate statistical test of the goodness of fit between hierarchy and data. However, it, too, has several disadvantages. Firstly, while in theory it subsumes the White and Clark test in practice the Dayton and Macready method is not yet sufficiently developed to allow this. Secondly, this results in the need to make mastery decisions for each subject with respect to each skill. Thirdly, the statistical methods applied by Dayton and Macready are not robust when even a small number of misclassifications occur with respect to 'true' mastery status. The appropriate classification of subjects as masters or non-masters is therefore of the utmost importance. Despite these limitations the Dayton and Macready method appears to be the most promising available test of the validity of an hierarchy. In addition to considering the hierarchy as a whole it also allows a direct test of the goodness of fit of alternative hierarchies to the same data.

The suggestion has been made that application of the White and

Clark test and the Dayton and Macready method to the same data may not lead to acceptance of the same 'validated' hierarchy. It is important to examine this suggestion in the light of empirical evidence. Except in the present study no such evidence exists. In the present study it was found that different 'validated' hierarchies emerged as a result of applying the two methods to the same data.

A third 'psychometric' method, the ordering-theoretic method, was also applied to the same data. The ordering-theoretic method, like the White and Clark test, considers skills in pairs. The method appears to be inferior to the White and Clark test in that it does not allow for errors of measurement, requires a mastery decision and does not include an appropriate statistical test. However, it is very simple to apply and may not yield results significantly different to those obtained from application of the White and Clark test. The relationship between the result of applying each method to the same data has not been tested except in the present data. The results obtained support the possibility of substituting the ordering-theoretic method for the White and Clark test although it must be recognized that the results obtained from one study may only be considered suggestive. A further consideration relating to which of these methods should be used relates to the purposes of the particular investigation. If the relationship between a particular pair(s) of skills is of major importance, the White and Clark test is the most appropriate available. If the validity of a composite hierarchy represents the central question the Dayton and Macready method appears to be most appropriate. In this case a preliminary sifting of the data to indicate the most likely hierarchy(ies) may be necessary. The

White and Clark test and the ordering-theoretic method appear to be appropriate for this purpose. However, if similar hierarchies emerge from the application of these two methods and these hierarchies are to be further tested by application of the Dayton and Macready method, the simpler ordering-theoretic method may be considered appropriate.

Two further implications relating to the methodology of learning hierarchy validation emerge from the present study. The first is that testing after instruction rather than during instruction is appropriate. The second is that it is possible to identify an hierarchy which is valid in terms of its transfer characteristics as well as its psychometric characteristics when instruction is delivered in a normal classroom setting rather than in a programmed format.

With respect to the sequencing of instruction leading to learning of the mole concept several implications of the present study may be stated. Firstly, a major implication is that a number of intellectual skills have been identified each of which is a necessary prerequisite to the learning of the key skill in the mole concept. Thus, the ability to determine the relative number of particles in given mole quantities of substances, the ability to inter-convert the number of moles of a substance and the number of particles present, the ability to determine the masses of different substances which contain the same number of moles, the ability to inter-convert masses of substances and the numbers of moles present, the ability to relate the meaning of the term mole to molar mass and the Avogadro number of particles, and finally to exhibit understanding of molar mass as a ratio are all skills which are subordinate to the ability to relate masses of substances in terms of the relative numbers of particles present. The actual arrangement of these skills is represented in Figure 15 (p. 191).

Several aspects of the validated hierarchy are of interest. First, there is dispute amongst teachers about whether there is a need for students to be able to calculate the actual number of particles in given mole and mass quantities of substances. The data in the present study support the need to develop this ability, and in turn therefore to gain an understanding of 'mole' in terms of the Avogadro number of particles. This relationship had been hypothesized to exist by the investigator on the grounds that it may be easier for the learner to work through an actual number of particles, even though the number is unusually large, than for him to think in terms of relative numbers. Another relationship which appears to depend upon psychological factors rather than upon logical relationships is that the learner needs to be able to calculate molar masses and to relate masses of different substances in terms of the numbers of moles present before he can determine the actual number of particles or relative number of particles present in masses of substances. This relationship was not hypothesized but is possibly related to the fact that an understanding of mass (or at least weight) is nearer the learner's direct experience than are relationships involving atoms. A further more tentative suggestion is that it may be necessary for the learner to be able to calculate molar masses before he can comprehend the formal meaning of the term mole. Finally, students have little difficulty in calculating molar masses of compounds. This appears to be further reflected in the lack of a consistent hierarchical relationship between the ability to apply the same fundamental skills to elements and compounds respectively. In general subjects who could apply the skills representing the mole concept to test items involving elements could also do so for test items relating to compounds.

The final implication of the present study relates to the relationship between learner developmental level and acquisition of the skills involved in learning the mole concept. It had been anticipated that correlations between developmental level test scores and scores on the tests of intellectual skills would be greater for the most complex skills. No pattern was observed. Further the actual values of these correlations were never more than moderate and were generally significantly smaller than the correlations between skills exhibiting an hierarchical relationship to one another. In terms of the proportion of variance in the scores for tests of the intellectual skills representing the mole concept which could be accounted for, the influence of the availability of subordinate skills was much greater than the influence of developmental level. This finding is supportive of the hierarchical model of learning, and tends to deny the importance of developmental level. However, several qualifications must be placed upon this interpretation. Firstly, the finding only relates to the particular content studied. Secondly, the tendency of individuals to revert to concrete thought when faced with novel problems tends to decrease their apparent developmental level. This would tend to decrease the correlations observed between these test scores and achievement scores. In addition the written nature of the responses to the developmental tests would be expected to decrease the precision with which subjects were classified as concrete or formal. Despite these qualifications the overwhelmingly greater influence of the availability of subordinate skills is impressive testimony to the power of the hierarchical model.

Suggestions for Further Research

A number of suggestions for further research may be made. These fall into three categories. The first relates to the methodology of learning hierarchy validation, the second to instruction in science, and the third to psychological considerations relating to science instruction. The following suggestions are made.

1. The relationship between the results of applying the White and Clark test, the ordering-theoretic method to the same data should be further investigated.

2. The Dayton and Macready model should be extended to allow the testing of larger hierarchies and to allow for unrestricted mis-classification parameters in situations additional to those involving the concept attainment model.

3. The White and Clark test and the Dayton and Macready method should be applied in studies relating to developmental hierarchies.

4. The application of the hierarchical model of learning should be applied to extension of the mole concept as defined in the present study. For example, the use of the mole in stoichiometric calculations should be investigated through this model.

5. It seems likely that many concepts in physical science are amenable to investigation through the application of the hierarchical model. This possibility should be investigated further in the light

of the more appropriate methods now available for learning hierarchy validation.

6. The relationship between developmental level and acquisition of intellectual skills should be investigated in a variety of content areas of different levels of sophistication.

BIBLIOGRAPHY

Airasian, P.W. & Bart, W.M. Validating a priori instructional hierarchies. *Journal of Educational Measurement*, 1975, 12(3), 163-173.

Allen, L.S. A scalogram analysis of classificatory behavior. *Journal of Research in Science Teaching*, 1970, 7(1), 43-45.

Ausubel, D.P. *Educational psychology: A cognitive view.* New York: Holt, Rinehart and Winston, Inc., 1968.

Bart, W.M. & Krus, D.J. An ordering-theoretic method to determine hierarchies among items. *Educational and Psychological Measurement*, 1973, 33, 291-300.

Bass, J.E. & Montague, E.J. Piaget based sequences of instruction. *Science Education*, 1972, 56(4), 503-512.

Beeson, E.W. Hierarchical learning in electrical science, *Journal of Research in Science Teaching*, 1977, 14(2), 117-127.

Beilin, H. The training and acquisition of logical operations. In *Piagetian cognitive-development and mathematical education*, M.F. Rosskopf (Ed.), National Council of Teachers of Mathematics, Washington, D.C., 1971.

Berliner, D.C. & Gage, N.L. The psychology of teaching methods. In Gage, N.L. (Ed.), The psychology of teaching methods: *The Seventy Fifth Yearbook of the National Society for the Study of Education.* Chicago, University of Chicago Press, 1976.

Boblick, J.M. Applying the systems approach to curriculum development in the science classroom. *Science Education*, 1971, 55(2), 103-113.

Brainerd, C.J. Neo-Piagetian training experiments revisited: Is there any support for the cognitive-developmental stage hypothesis. *Cognition*, 1973, 2(3), 349-370.

Bredderman, T.A. The ability to combine and control variables. *Science Education*, 1974, 58, 457-469.

Brennan, R.L. *The evaluation of mastery test items.* (ERIC Document 092 593), 1974.

Briggs, L.J. *Sequencing of instruction in relation to hierarchies of competence.* Pittsburgh: American Institute for Research, 1968.

Bruner, J.S. *The process of education.* Cambridge, Massachusetts: Harvard University Press, 1960.

Bruner, J. On conservation of liquids, in Bruner, Olver and Greenfield (Eds.), *Studies in cognitive growth.* New York: Wiley, 1966, 183-207.

Capie, W. & Jones, H.W. An assessment of hierarchy validation techniques. *Journal of Research in Science Teaching*, 1971, 8, 137-147.

Carroll J.B. Discussion. In Resnick, L.B. (Ed.), *Hierarchies in children's learning. Instructional Science*, 1973 (2), 311-362.

Chiappeta, E.L. A review of Piagetian studies relevant to science instruction at the secondary and college level. *Science Education*, 1976, 60(2), 253-261.

Cohen, J. *A coefficient of agreement for nominal scales.* Educational and Psychological Measurement, 1960, 20, 37-46.

Cornish, R. A computer program for the generalized White and Clark test (personal communication).

Cotton, J.W., Gallagher, J.P., & Marshall, S.P. The identification and decomposition of hierarchical tasks. *American Educational Research Journal*, 1977, 14(3), 189-212.

Dayton, C.M. & Macready, G.B. A probabilistic model for validation of behavioral hierarchies. *Psychometrika*, 1976a, 41(2), 189-204.

Dayton, C.M. & Macready, G.B. Computer programs for probabilistic models. University of Maryland, 1976b.

Dence, J.B. The mathematics needed in freshman college chemistry. *Science Education*, 1970, 287-290.

Denny, R.T. The mathematics skills test (MAST) for chemistry. *Journal of Chemical Education*, 1971 (12), 845-846.

De Rose, J.V. New directions for chemical education in high schools. In *The 1969 Star Awards.* Washington: National Science Teachers Association, 1969.

Dewey, J. *Democracy and education.* New York: Macmillan, 1916.

Duncan, I.M. & Johnstone, A.H. The mole concept. *Education in Chemistry*, 1973, 10, 213-214.

Eisenberg, T.A. & Walbesser, H. Learning hierarchies--numerical considerations. *Journal for Research in Mathematics Education*, 1971, 2(4), 244-256.

Flavell, J.H. *The developmental psychology of Jean Piaget.* New York: Van Nostrand Co., 1963.

Gagné, R.M. The acquisition of knowledge. *Psychological Review*, 1962, 69(4), 355-365.

Gagné, R.M. A psychologist's counsel on curriculum design. *Journal of Research in Science Teaching*, 1963, 1(1), 27-32.

Gagné, R.M. *The conditions of learning* (1st ed.). New York: Holt, Rinehart and Winston, 1965.

Gagné, R.M. Curriculum research and the promotion of learning. In Tyler, R.W., Gagné, R.M., & Scriven, M. (Eds.), *Perspectives of Curriculum Evaluation.* Chicago: Rand McNally, 1967.

Gagné, R.M. Contributions of learning to human development. *Psychological Review*, 1968, 75(3), 177-191.

Gagné, R.M. *Some notes on criterion-referenced measurement.* Unpublished manuscript, Florida State University, 1969a.

Gagné, R.M. The acquisition of knowledge. *Psychological Review*, 1969b, 69(4), 355-365.

Gagné, R.M. *The conditions of learning* (2nd ed.). New York: Holt, Rinehart and Winston, 1970a.

Gagné, R.M. Some new views of learning and instruction. *Phi Delta Kappan*, 1970b, 51(9), 468-472.

Gagné, R.M. Domains of learning. *Interchange*, 1972, 3, 1-8.

Gagné, R.M. Learning and instructional sequence. In F. Kerlinger (Ed.), *Review of Research in Education (1).* Itasca, Illinois: F.E. Peacock, 1973.

Gagné, R.M. *Essentials of learning for instruction.* Hinsdale, Illinois: The Dryden Press, 1974.

Gagné, R.M. *The conditions of learning* (3rd ed.). New York: Holt, Rinehart and Winston, 1977.

Gagné, R.M. & Paradise, M.E. [illegible]ilities and learning sets in knowledge acquisition. *Psychological Monographs*, 1961, 75 (whole No. 518).

Gagné, R.M., Mayor, J.R., Garstens, H.L., & Paradise, N.E. Factors in acquiring knowledge of a mathematical task. *Psychological Monographs* (whole No. 526), 1962.

Gagné, R.M. & Briggs, L.J. *Principles of instructional design.* New York: Holt, Rinehart and Winston, 1974.

Gardner, M. (Ed.). *Interdisciplinary approaches to chemistry.* New York: Harper and Row, 1973.

Glaser, R. & Nitko, A.J. Measurement in learning and instruction. In Thorndike, R.L. (Ed.), *Educational measurement* (end ed.). Washington, D.C.: American Council on Education, 1971.

Glaser, R. & Resnick, L. Instructional psychology. *Annual Review of Psychology*, 1973, 23, 181-276.

Good, R. & Morris, G. *Mathematics and logic skills exhibited by college freshman chemistry students.* A paper presented to NARST, Toronto, 1978.

Gower, D.M., Daniels, D.J. & Lloyd, G. The mole concept. *School Science Review*, 1977a, 58, 658-676.

Gower, D.M., Daniels, D.J. & Lloyd, G. Hierarchies among the concepts which underlie the mole. *School Science Review*, 1977b, 59, 285-299.

Griffiths, A.K. & Cornish, A.G. *An analysis of three recent methods for the identification and validation of learning hierarchies.* A paper presented at AERA, Toronto, 1978.

Guttman, L.A. Basis for scaling qualitative data. *American Sociological Review*, 1944, 9, 139-150.

Haladyna, T.M. Effects of different samples on items and test characteristics of criterion-referenced tests. *Journal of Educational Measurement*, 1974, 11, 93-99.

Hambleton, R.K. & Gorth, W.P. *Criterion-referenced testing: Issues and application.* Amherst, Massachusetts: University of Massachusetts, 1971. (ERIC Document ED 060 025).

Hambleton, R.K. & Novick, M.R. Toward an integration of theory and method for criterion-referenced tests. *Journal of Educational Measurement*, 1973, 10, 159-170.

Hambleton, R.K. & others. *Criterion-referenced testing and measurement: A review of technical issues and developments.* Invited symposium presented at the annual meeting of AERA, Washington, D.C., 1975.

Harris, C.W. An interpretation of Livingston's reliability coefficient for criterion-referenced tests. *Journal of Educational Measurement*, 1972, 9, 27-29.

Harrison, D.B. *Reflective intelligence and mathematics learning.* Unpublished doctoral dissertation, University of Alberta, 1967.

Head, J.O. Teaching the mole concept in schools. *School Science Review*, 1968, 924-925.

Herron, J.D. Piaget for chemists. *Journal of Chemical Education*, 1975, 52(3), 146-150.

Herron, J.D. *What research says to the college science teacher.* A paper presented at NSTA, Philadelphia, 1976a.

Herron, J.D. Commentary on Piagetian cognitive development and achievement in science. *Journal of Research in Science Teaching*, 1976b, 13(4), 355-357.

Herron, J.D., Cantu, L.L., Ward, R. & Srinivasan, V. Problems associated with concept analysis. *Science Education*, 1977, 61(2), 185-200.

Heslop, R.B. & Wild, G.M. *S.I. units in chemistry*. London: Applied Science Publications Ltd., 1975.

Hively, W., Patterson, H.L., & Page, S.A. A "universe-defined" system of arithmetic achievement tests. *Journal of Educational Measurement*, 1968, 5, 275-290.

Hobbs, E.T. *Formal operations in secondary students.* Unpublished doctoral dissertation, University of Alberta, 1975.

Howe, A. *Formal operational thought and the high school science curriculum.* A paper presented at NARST, Chicago, 1974.

Hurd, P. de Hart. *New directions in teaching secondary school science.* Chicago: Rand McNally, 1969.

Ingle, R.B. & Shayer, M. Conceptual demands in Nuffield 'O' level chemistry. *Education in Chemistry*, 1970, 7, 182-183.

Inhelder, B. & Piaget, J. *The growth of logical thinking.* USA: Basic Books, 1958.

Karplus, Robert. Science teaching and the development of reasoning. *Journal of Research in Science Teaching*, 1977, 14(2), 169-175.

Karplus, E.F. & Karplus, R. Intellectual development beyond elementary school. *School Science and Mathematics*, 1970, 70, 398-406.

Karplus, R. & Lavatelli, C. *The developmental theory of Piaget: Conservation.* San Francisco: Davidson Films, 1969.

Karplus, E., Karplus, R. & Wollman, W. Intellectual development beyond elementary school (IV). *School Science and Mathematics*, 1974, 74, 476-482.

Karplus, R., Karplus, E.F., Formisano, M., & Paulsen, A.C. *Proportional reasoning and control of variables in seven countries.* Berkeley, California: Lawrence Hall of Science, 1975.

Kass, H. *Structure of perceived relations among physics concepts.* Unpublished doctoral dissertation, University of Alberta, 1969.

Kofsky, E. A scalogram study of classificatory development. *Child Development*, 1966, 37(1), 191-204.

Kolb, J.R. Effects of relating mathematics to science instruction on the acquisition of quantitative science behavior. *Journal of Research in Science Teaching*, 1967, 5, 174-182.

Lawson, A.E. The development and validation of a classroom test of formal reasoning. *Journal of Research in Science Teaching*, 1978, 15(1), 11-24.

Lawson, A.E. & Nordland, F.H. Conservation reasoning ability and performance on BSCS blue version examinations. *Journal of Research in Science Teaching*, 1977, 14(1), 69-76.

Lawson, A.E. & Renner, J.W. Quantitative analysis of responses to Piagetian tasks and its implications for curriculum. *Science Education*, 1974, 58(4), 454-559.

Lawson, A.E. & Renner, J.W. Relationships of science subject matter and developmental levels of learners. *Journal of Research in Science Teaching*, 1975, 12(4), 347-358.

Light, R.L. Issues in the analysis of qualitative data. In R.M.W. Travers (Ed.), *Second Handbook of Research on Teaching.* Chicago: Rand McNally, 1973.

Lingoes, J.C. Multiple scalogram analysis: A set-theoretic model for analyzing dichotomous items. *Educational and Psychological Measurement*, 1963, 23(3), 501-524.

Linke, R.D. Replicative studies in hierarchical learning of graphical interpretation skills. *British Journal of Educational Psychology*, 1975, 45, 39-46.

Livingston, S.A. Criterion-referenced applications of classical test theory. *Journal of Educational Measurement*, 1972, 9(1), 13-26.

Lunzer, E. Problems of formal reasoning in test situations. In *Cognitive development in children.* Chicago: University of Chicago Press, 1965.

Lunzer, E. Address to the Jean Piaget Society. Philadelphia, 1973.

Munby, H. *Piagetian research in science education: Some misgivings about its potential to improve practice.* A paper presented at NARST, Toronto, 1978.

Newstead, B. Mole relationships. *Chem 13 News*, 1978, 93, 11.

Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K. & Bent, D.H. *Statistical package for the social sciences* (2nd ed.). New York: McGraw-Hill, 1975.

Novak, J.D. *A theory of education.* Ithaca, New York: Cornell University Press, 1977.

Novick, S. & Menis, J. A study of student perceptions of the mole concept. *Journal of Chemical Education*, 1976, 11, 720, 722.

Nuffield 'O' level chemistry: Teacher's guide. England: Longmans, 1967.

Oakland, T. *Evaluation of models for estimating reliability and validity of criterion-referenced measures*, 1972. (ERIC Document ED 065 589).

O'Connor, P.R. & others. *Chemistry, experiments and principles, teacher's guide.* Lexington, Massachusetts: D.C. Heath and Co., 1968.

Okey, J.R. & Gagné, R.M. Revision of a science topic using evidence of performance on subordinate skills. *Journal of Research in Science Teaching*, 1970, 7, 321-325.

Okey, J.R. & Gagné, R.M. Revision of a science topic using evidence of performance on subordinate skills. In Gagné, R.M., *Basic studies of learning hierarchies in school subjects*, 1971. (ERIC Document Ed 039 611).

Parry, R.W. *et al. Chemistry: Experimental foundations: Teacher's guide.* London: Prentice-Hall, 1969.

Phillips, D.C. & Kelly, M.E. Hierarchical theories of development in education and psychology. *Harvard Educational Review*, 1975, 45, 351-375.

Phillips, D.G. The development of the concept of displacement volume: A hierarchical model and its partial testing under two methods of presentation. *Journal of Research in Science Teaching*, 1971, 8(1), 9-19.

Phillips, E.R. *Development of optimal instructional sequences.* A paper presented at AERA, Chicago, 1974. (ERIC Document 097 374).

Piaget, J. *The origins of intelligence in children.* New York: International University Press, 1952.

Piaget, J. Cognitive development in children. *Journal of Research in Science Teaching*, 1964, 2, 170-186.

Piaget, J., Inhelder, B., & Szeminska, A. *The child's conception of geometry*. New York: Basic Books, 1960.

Pimentel, G.C. (Ed.). *Chemistry: An experimental science*. San Francisco: W.H. Freeman, 1963.

Popham, W.J. & Husek, T.R. Implications of criterion referenced measurement. *Journal of Educational Measurement*, 1969, 6(1), 1-9.

Posner, G.J. & Strike, K.A. A categorization scheme for principles of sequencing content. *Review of Educational Research*, 1976, 46(4), 665-690.

Proctor, C.H. A probabilistic formulation and statistical analysis for Guttman scaling. *Psychometrika*, 1970, 35, 73-78.

Raven, R. The development of the concept of momentum in primary school children. *Journal of Research in Science Teaching*, 1968, 5, 216-223.

Raven, R.J. The development of the concept of acceleration in elementary school children. *Journal of Research in Science Teaching*, 1972, 9(3), 201-206.

Raven, R.J. The development of a test of Piaget's logical operations. *Science Education*, 1973, 3, 377-385.

Renner, J.W. *Evaluating intellectual development using written responses to selected science problems*. Norman, Oklahoma: University of Oklahoma, 1977.

Renner, J.W. & Stafford, D.G. *Teaching science in the secondary school*. New York: Harper and Row, 1972.

Resnick, L.B. Issues in the study of learning hierarchies. In Resnick, L.B. (Ed.), Hierarchies in children's learning. *Instructional Science*, 1973 (2), 311-362.

Resnick, L.B. & Wang, M.C. *Approaches to the validation of learning hierarchies*. Educational Testing Service, 1969.

Roudabush, G.E. & Green, D.R. *Aspects of a methodology for creating criterion-referenced tests*. A paper presented at the annual meeting of the National Council for Measurement in Education, Chicago, 1972, (ERIC Document ED 099 411).

Rowell, J.A. & Hoffman, P.J. Group tests for distinguishing formal operational thinkers. *Journal of Research in Science Teaching*, 1975, 12(2), 157-164.

Ruda, P.T. Another mole triangle. *Chem 13 News*, 1978, 94, 15.

Sayre, S. & Ball, D.W. Piagetian cognitive development and achievement in science. *Journal of Research in Science Teaching*, 1975, 12, 165-174.

Schwab, J.J. The teaching of science as enquiry. In *The teaching of science* (Schwab, J.J. and Brandwein, P.F.). Cambridge, Massachusetts: Harvard University Press, 1966.

Science--A Process Approach. American Association for the Advancement of Science, Xerox Corporation, 1968.

Seddon, G.M. A comparison of three different measures for predicting achievement in chemistry in the age range 15 to 18 plus. *Research in Education*, 1974, 12, 63-70.

Shayer, M. Chemistry for the sixteen-year-old school leaver. In *The discipline of chemistry--Its place in education*. London: The Chemical Society, 1973.

Shayer, M. & Wharry, D. Piaget in the classroom (Part 1): Testing a whole class at the same time. *School Science Review*, 1974, 55, 447-458.

Shulman, L.S. Psychological controversies in the teaching of science and mathematics. *The Science Teacher*, 1968, 35(6), 34-38.

Shulman, L.S. The psychology of school subjects: A premature obituary? *Journal of Research in Science Teaching*, 1974, 11(4), 319-339.

Shulman, L.S. & Tamir, P. Research on teaching in the natural sciences. In Travers, R.M. (Ed.), *Second Handbrook of Research on Teaching*. Chicago: Rand McNally, 1973.

Skemp, R.H. Reflective intelligence and mathematics. *British Journal of Educational Psychology*, 1961, 31, 4 -55.

Smith, P.J. Piaget in high school science instruction. *Journal of Chemical Education*, 1978, 55(2), 115-118.

Sticht, R. Comments on Lovell's paper: Does learning recapitulate ontogeny? In Green, D.R., Ford, M.P., & Flamer, G.B. (Eds.), *Measurement and Piaget*. New York: McGraw-Hill, 1971.

Strauss, S. Learning theories of Gagné and Piaget: Implications for curriculum development. *Teachers College Record*, 1972, 74(1), 81-102.

Swaminathan, H., Hambleton, R.K., & Algina, J. Reliability of criterion-referenced tests: A decision-theoretic formulation. *Journal of Educational Measurement*, 1974, 11(4), 261-267.

Walbesser, H.H. & Eisenberg, T.A. *A review of research on behavioral objectives and learning hierarchies*. (ERIC Document 059 900), 1972.

Wellens, B., Lenke, J.M. & Oswald, J.H. *An investigation of a proposed hierarchy of mathematics concepts and skills using ordering theory.* A paper presented at AERA, New York, 1977.

Wheeler, A.E. *Student misconceptions in chemical equilibrium as related to achievement and cognitive level.* Unpublished M.Ed. thesis, The University of Alberta, 1973.

Wheeler, A.E. *Proportionality in high school chemistry.* Unpublished doctoral dissertation, University of Alberta, 1976.

Wheeler, A.E. & Kass, H. *Proportional reasoning in introductory high school chemistry.* A paper presented to NARST, Cincinnati, Ohio, 1977.

Wheeler, A.E. & Kass, H. Student misconceptions in chemical equilibrium. *Science Education*, 1978, 62(2), 223-232.

White, R.T. Research into learning hierarchies. *Review of Educational Research*, 1973, 43(3), 361-375.

White, R.T. A model for validation of learning hierarchies. *Journal of Research in Science Teaching*, 1974a, 11(1), 1-3.

White, R.T. Indexes used in testing the validity of learning hierarchies. *Journal of Research in Science Teaching*, 1974b, 11(1), 61-66.

White, R.T. & Clark, R.C. A test of inclusion which allows errors of measurement. *Psychometrika*, 1973, 38, 77-86.

Wiegand, V.K. *A study of subordinate skills in science problem solving.* Doctoral dissertation, University of California, Berkeley, 1969. (Dissertation Abstracts International, 1970, 31, 690A).

Wilson, R.J. & Burnett, J.D. *Estimating the reliability of criterion-referenced tests.* A paper presented at the annual conference of the Canadian Society for the Study of Education, Edmonton, June, 1975.

Wood, W. *A significance test for ordering theory.* Doctoral dissertation, Boston College, 1974. (Dissertation Abstracts International, 1975, 35, 7145A).

Woodson, M.I.C.E. The issue of item and test variance for criterion-referenced tests. *Journal of Educational Measurement*, 1974, 11, 63-64.

APPENDICES

APPENDIX 1

Quiz 1 Closed Book NAME ____________________

Attempt all questions. For questions one to eight circle the correct answer.

You are given the following atomic weights:

carbon	- 12	oxygen	- 16
chlorine	- 35.5	phosphorus	- 31
hydrogen	- 1	sodium	- 23
nitrogen	- 14	sulfur	- 32

The Avogadro number = 6.02×10^{23}

If you have any idea how to tackle a question, even if you cannot complete it, DO WHAT YOU CAN.

1. Which of the following, if any, best describes the meaning of the term 'atomic weight?'
   (a) the weight of one atom of the element.
   (b) the weight of one million atoms of the element.
   (c) the weight of one atom of the element relative to the weight of one atom of another element taken as agreed standard by scientists.
   (d) cannot be determined with certainty because the atom is too small.
2. If the atomic weight of magnesium is 24, what is the atomic weight of an element whose atoms are 6 times heavier?

(a) 6 (b) 24

(c) 144 (d) cannot answer, not enough information

3. Which of the following represents one mole of copper (atomic weight 64)?

(a) 64 atoms (b) 10^6 atoms

(c) $\frac{6.02 \times 10^{23}}{64}$ atoms (d) 6.02×10^{23} atoms

4. Which of the following represents one mole of calcium (atomic weight 40)?

(a) $\frac{40}{12}$ grams (b) 40 grams

(c) $\frac{6.02 \times 10^{23}}{40}$ grams (d) $40 \times 6.02 \times 10^{23}$ grams

5. Which of the following best describes the meaning of the term 'molecular weight' of a compound?

(a) the weight of one molecule of the compound.

(b) the weight of one million (10^6) molecules of the compound.

(c) the weight of 22.4 molecules of the compound.

(d) the weight of one molecule relative to the weight of one atom of an element taken as standard for the atomic weight scale.

6. If the molecular weight of a gas is 44, what is the molecular weight of a gas whose molecules are twice as heavy?

(a) 2 (b) 22

(c) 88 (c) some other number

7. Which of the following represents one mole of nitric oxide (NO) (molecular weight 30)?

(a) 30 grams

(b) $\frac{30}{12}$ grams

(c) $\frac{6.02 \times 10^{23}}{30}$ grams

(d) $30 \times 6.02 \times 10^{23}$ grams

8. Which of the following represents one mole of hydrogen sulfide (H_2S, molecular weight 34)?

(a) 34 molecules

(b) 6.02×10^{23} molecules

(c) 10^6 molecules

(d) $\frac{6.02 \times 10^{23}}{34}$ molecules

9. Calculate the weight of one mole of:

(a) hydrogen chloride (HCl)

(b) carbon monoxide (CO)

(c) nitrogen dioxide (NO_2)

(d) phosphine (PH_3)

(e) phosphorus trioxide (P_2O_3)

(f) sodium sulfide (Na_2S)

APPENDIX 2

Quiz 2 Closed Book NAME ______________________

Answer all questions. Please show all your work. You will lose no marks for any working which is incorrect, only for incorrect or incomplete answers.

You are given the following atomic weights:

carbon	- 12	nitrogen	- 14
calcium	- 40	oxygen	- 16
hydrogen	- 1	potassium	- 39
iron	- 56	sulfur	- 32

The Avagadro number = 6.02×10^{23}

DO NOT SPEND TOO LONG ON ANY QUESTION YOU HAVE DIFFICULTY ANSWERING.

If you have any idea how to tackle a question, even if you cannot complete it, DO WHAT YOU CAN.

1. How many atoms are there in:

   (a) 2 moles carbon

   (b) 5 moles iron

2. How many moles are there in:

   (a) 6.02×10^{23} atoms of iron

(b) 3.01×10^{23} atoms of carbon

3. How many molecules are there in:

(a) 2 moles carbon dioxide (CO_2)

(b) one-tenth of a mole of hydrogen sulfide (H_2S)

4. How many moles are there in:

(a) 6.02×10^{23} molecules of carbon dioxide (CO_2)

(b) 3.01×10^{25} molecules of sulphur dioxide (SO_2)

5. How many moles are represented by:

(a) 24g carbon

(b) 10g calcium

6. What is the weight of:

(a) 2 moles sulfur

(b) 2 moles of potassium

7. How many moles are represented by:

   (a) 60g nitric oxide (NO)

   (b) 23g nitrogen dioxide (NO_2)

8. What is the weight of:

   (a) 2 moles of carbon dioxide (CO_2)

   (b) one-tenth of a mole of hydrogen sulfide (H_2S)

9. How many moles of potassium contain the same number of atoms as 0.5 moles calcium?

10. How many atoms of sulfur are represented by the number of moles containing 10^{10} atoms of carbon?

11. How many moles of carbon dioxide (CO_2) contains the same number of molecules as 1.5 moles of water (H_2O)?

12. How many molecules of sulfur dioxide (SO_2) are represented by the number of moles containing 10^{20} molecules of water (H_2O)?

13. Calculate the mass of carbon representing the same number of moles as 64 grams of sulfur.

14. How many moles of iron would weigh the same as 2 moles of carbon?

15. Calculate the mass of water (H_2O) containing the same number of moles as 34 grams of hydrogen sulfide (H_2S).

16. How many moles of carbon dioxide (CO_2) would weigh the same as 5 moles of water (H_2O)?

APPENDIX 3

Quiz 3   Closed Book   NAME ____________________

Answer all questions. Please show all your work. You will lose no marks for any working which is incorrect, only for incorrect or incomplete answers.

You are given the following atomic weights:

aluminum	27	nitrogen	14
carbon	12	oxygen	16
copper	64	potassium	39
hydrogen	1	sulfur	32

The Avogadro number = 6.02×10^{23}

DO NOT SPEND TOO LONG ON ANY QUESTION YOU HAVE DIFFICULTY ANSWERING.

If you have any idea how to tackle a question, even if you cannot complete it, DO WHAT YOU CAN.

1. How many atoms are there in 6.4 grams of copper?

2. What is the mass of 3.01×10^{25} atoms of sulfur?

3. Calculate the mass of potassium containing the same number of atoms as 108 grams of aluminum.

4. Calculate the mass of carbon containing 4 times as many atoms as 64 grams of sulfur.

5. How many molecules are there in 8.8 grams of carbon dioxide (CO_2)?

6. What is the mass of 6.02×10^{22} molecules of sulfur dioxide?

7. Calculate the mass of hydrogen sulfide (H_2S) containing the same number of molecules as 54 grams of water (H_2O).

8. Calculate the mass of carbon monoxide (CO) containing 3 times as many molecules as 90 grams of nitric oxide (NO).

APPENDIX 4A

NAME ____________________

DATE ____________________

Chemistry 10 Test

This is a closed book test. Attempt all questions. Answer multiple choice questions by circling the appropriate letter.

In the questions which are not multiple choice, it is very important that you show all your work. You will lose no marks for any work which is incorrect, only for incorrect or incomplete answers.

Please use the following atomic weights:

aluminum	- 27	magnesium	- 24
calcium	- 40	nitrogen	- 14
carbon	- 12	oxygen	- 16
chlorine	- 36.5	potassium	- 39
copper	- 64	silver	- 108
hydrogen	- 1	sodium	- 23
iron	- 56	sulfur	- 32

The Avogadro number (N) is 6.02×10^{23}

DO NOT SPEND TOO LONG ON ANY ONE QUESTION WHICH YOU HAVE DIFFICULTY ANSWERING.

If you have any idea how to tackle a question, even if you cannot complete it, DO WHAT YOU CAN.

1. How many moles of magnesium would weigh the same as six moles of carbon?

2. How many molecules are there in 23 grams of nitrogen dioxide (NO_2)?

3. How many moles of the named compound are there in:

   (a) 6.02×10^{23} molecules of carbon monoxide (CO)?

   (b) 3.01×10^{25} molecules of sulfur trioxide (SO_3)?

4. Consider 10^{12} atoms of calcium. How many atoms of carbon are needed to give the same number of moles as this number of atoms of calcium?

5. Which of the following represents one mole of hydrogen chloride (HCl)?

   (a) $\frac{36.5}{12}$ grams (b) 36.5 grams

   (c) $\frac{6.02 \times 10^{23}}{36.5}$ grams (d) $36.5 \times 6.02 \times 10^{23}$ grams

6. Which of the following, if any, describes the meaning of atomic weight?

(a) the weight of one atom of the element.

(b) the weight of one million atoms of the element.

(c) the weight of one atom of the element relative to the weight of one atom of another element taken as a standard agreed upon by scientists.

(d) cannot be determined with certainty because the atom is too small.

7. How many atoms are there in:

(a) 2 moles of copper? (b) half a mole of iron?

8. How many moles of nitrogen (N_2) contain the same number of molecules as 2 moles of carbon monoxide (CO)?

9. How many molecules are there in:

(a) 3 moles of hydrogen gas (H_2)? (b) one-tenth of a mole of hydrogen chloride (HCl)?

10. How many atoms are there in 10.8 grams of silver?

11. If the molecular weight of a gas is 33, what is the molecular weight of a gas whose molecules are three times as heavy?

(a) 3 (b) 11

(c) 99 (d) some other number

12. Calculate the weight of one mole of:

(a) sodium chloride (NaCl) (b) nitric oxide (NO)

(c) sulfur dioxide (SO_2) (d) methane (CH_4)

(e) ethane (C_2H_6) (f) hydrogen sulfide (H_2S)

13. Which of the following represents one mole of sodium?

(a) $\frac{23}{12}$ grams (b) 23 grams

(c) $\frac{6.02 \times 10^{23}}{23}$ grams (d) $23 \times 6.02 \times 10^{23}$ grams

14. How many moles of sodium contains the same number of atoms as 0.5 moles of calcium?

15. Calculate the mass of copper representing the same number of moles as 24 grams of carbon.

16. Which of the following best represents the meaning of the term "molecular weight" of a compound?

(a) the sum of the actual weights of the atoms in the molecule.

(b) the weight of one molecule of the compound relative to the weight of one molecule of hydrogen.

(c) the weight of 22.4 molecules of the compound.

(d) the weight of one molecule relative to the weight of one atom of the element taken as standard for the atomic weight scale.

17. What is the mass of 3.01×10^{25} atoms of carbon?

18. How many moles of the named compound are represented by:

(a) 88 grams of carbon dioxide (CO_2)?

(b) 32 grams of sulfur dioxide (SO_2)?

19. Calculate the mass of ammonia (NH_3) containing twice as many molecules as 32 grams of oxygen (O_2).

20. Which of the following represents one mole of magnesium?

(a) 24 atoms (b) 10^6 atoms

(c) $\frac{6.02 \times 10^{23}}{24}$ atoms (d) 6.02×10^{23} atoms

21. How many moles are there in:

(a) 6.02×10^{24} atoms of silver? (b) 3.01×10^{23} atoms of sulfur?

22. What is the weight of:

(a) 2 moles of carbon monoxide (CO)?

(b) one-tenth of a mole of ammonia (NH_3)?

23. Calculate the mass of hydrogen (H_2) containing the same number of molecules as 14 grams of carbon monoxide (CO).

24. What is the mass of 6.02×10^{25} molecules of carbon dioxide (CO_2)?

25. Calculate the mass of ammonia (NH_3) representing the same number of moles as 28 grams of carbon monoxide (CO).

26. If the atomic weight of copper is 64, what is the atomic weight of an element whose atoms are four times as heavy?

(a) 4 (b) 16

(c) 256 (d) cannot answer, because not enough information is given

27. How many moles are represented by:

(a) 20 grams of calcium? (b) 69 grams of sodium?

28. Consider 10^6 molecules of hydrogen (H_2). How many molecules of carbon dioxide are needed to give the same number moles as this number of molecules of hydrogen?

29. Calculate the mass of calcium containing the same number of atoms as 108 grams of aluminum.

30. What is the weight of:

(a) 3 moles of calcium? (b) one-tenth of a mole of calcium?

31. How many moles of carbon monoxide (CO) would weigh the same as 4 moles of nitrogen (N_2)?

32. Calculate the mass of magnesium containing 3 times as many atoms as 36 grams of carbon.

33. Which of the following represents one mole of carbon monoxide (CO)?

(a) 28 molecules (b) 10^6 molecules

(c) $\frac{6.02 \times 10^{23}}{28}$ molecules (d) 6.02×10^{23} molecules

APPENDIX 4B

ITEMS TESTING SKILLS A TO H, FINAL TEST

A	(mass: relative # particles)	19, 23, 29, 32
B	(mass: # particles)	2, 10, 17, 24
C	(mole: relative # particles)	4, 8, 14, 28
D	(mole: # particles)	3, 7, 9, 21
E	(mass: equal # moles)	1, 15, 25, 31
F	(mass: # moles)	18, 22, 27, 30
G	(mole: molar mass, L particles)	5, 13, 20, 33
H	(molar mass as a ratio)	6*, 11, 16*, 26

* These items were dropped from the analysis.

N.B. Item 12 was included as a linking item between elements and compounds.

APPENDIX 5

THE TEST OF DEVELOPMENTAL LEVEL

NAME________________________

DATE________________________

This booklet consists of four exercises. IT IS NOT A TEST. However, we would like to see the approach you use in arriving at your solution to each task. It is therefore essential that you show all your work clearly.

The Two Solids (TS)

This task involves separately submerging two solids in some water in a graduate. Read the instructions carefully. DO NOT answer any questions until you are ready, BUT once you have begun a new question DO NOT change any previous answers. Answer the questions in the order in which they appear.

1. Compare the two [illegible]g., Are they the same colour, how do their weights [illegible]e they the same size, . . .?

2. Write down the number on the graduate which corresponds to the water level.

3. Very carefully lower solid A into the water until it rests on the bottom. Write down the new water level.

4. WITHOUT DOING IT, predict what water level you would expect if solid B had been put in the water instead of solid A.

5. Explain as carefully as you can how you arrived at your answer to question 4.

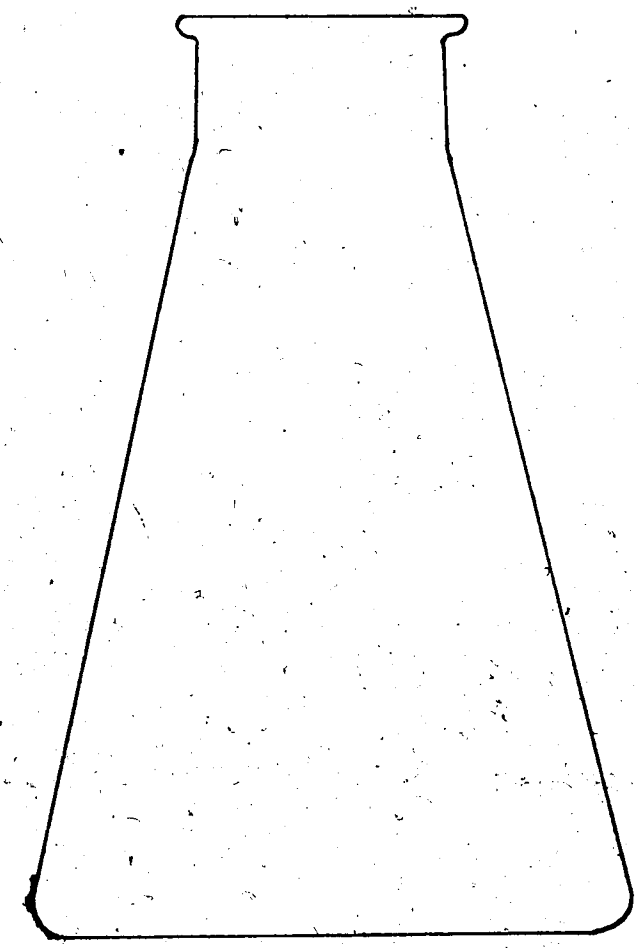
6. Carefully remove solid A from the graduate, and replace it with solid B. Write down the new water level.

7. Explain any differences in your answers to 4 and 6.

The Ratio Task (RT)

The figure drawn below represents a SMALL chemical flask. When measured with large paper clips, the height of the small flask was found to be 4 clips. When a similar LARGE flask (not shown in the diagram) was measured with the same large paper clips, it was found to be 6 clips high.

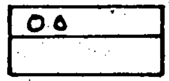
Now please do these things:

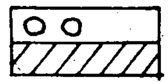
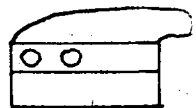
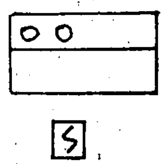
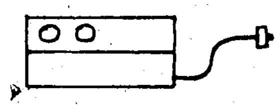


1. Measure the height of the small flask on the right using the SMALL paper clips provided. The height is _____ SMALL paper clips.
2. PREDICT the height of the LARGE flask if it were measured with the same small clips. The height of the LARGE flask would be _____.
3. EXPLAIN how you arrived at your prediction. You may use diagrams, words, or calculations. Please explain your steps carefully.

EXPLANATION (Be precise)

The Radio Task (RA)

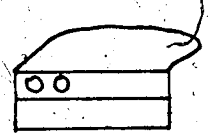
A transistor radio  costs $40, but you can buy any of the following "extras" if you want to:

	leather case		$10
OR	carrying strap		$5
OR	extra battery		$2
OR	ear phone		$1

Suppose you want to buy a radio. Write down or draw as many different choices as you can think of.

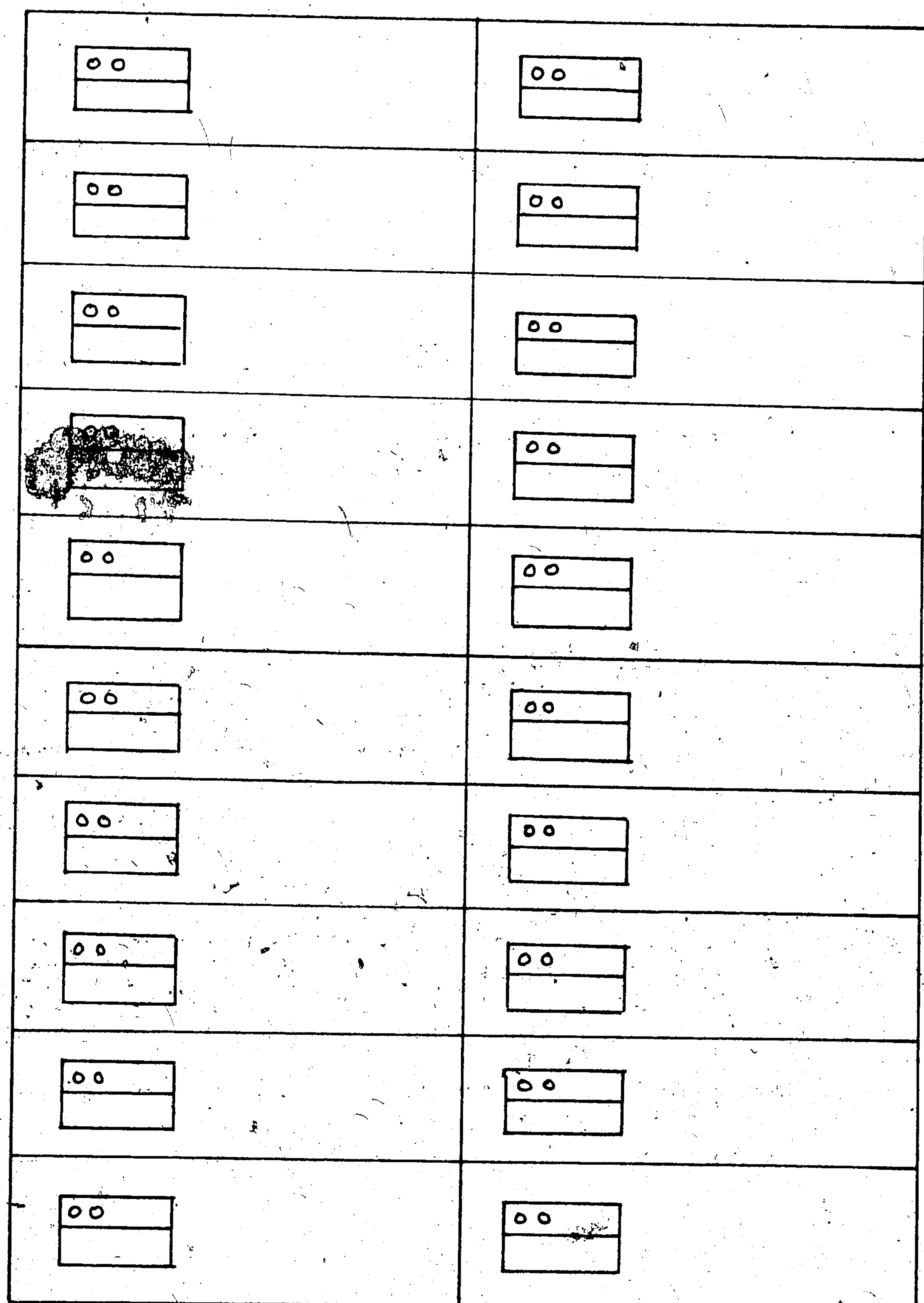
For example

radio + case + earphone



radio + strap

Put your answers on the special answer sheet on the next page.

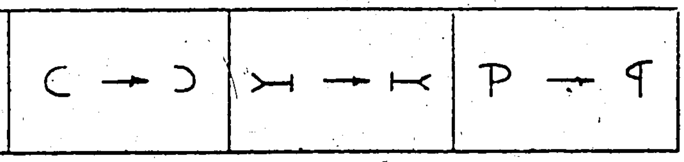


APPENDIX 6

THE SKEMP TEST

SK6: PRACTICE SHEET

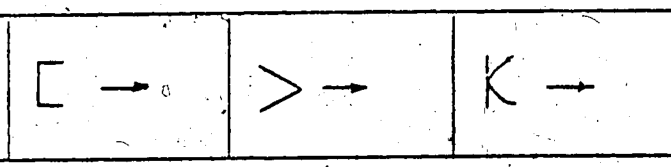
Operation 1



In the above figures, the one on the left of each pair has been changed to the one on the right by means of the same simple operation. In other words, the above figures given three examples of a particular operation. You have to find out what the operation is, and then do the same operation to some other figures.

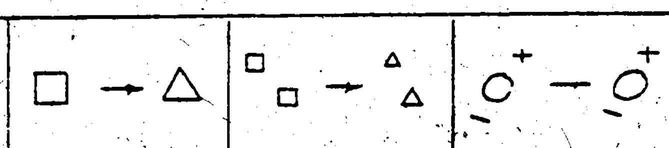
What is the operation? It is reversing from left to right. Do this on each of the figures below, and fill in the answers in the blank spaces. Check with the answers on the blackboard to make sure that you have understood.

Do Operation 1 on these.



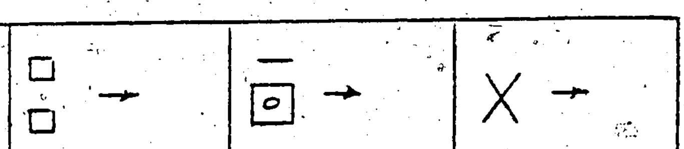
Here is a different operation:

Operation 2



When you have found out what it is, do it on the figures below. Check with the answers on the board.

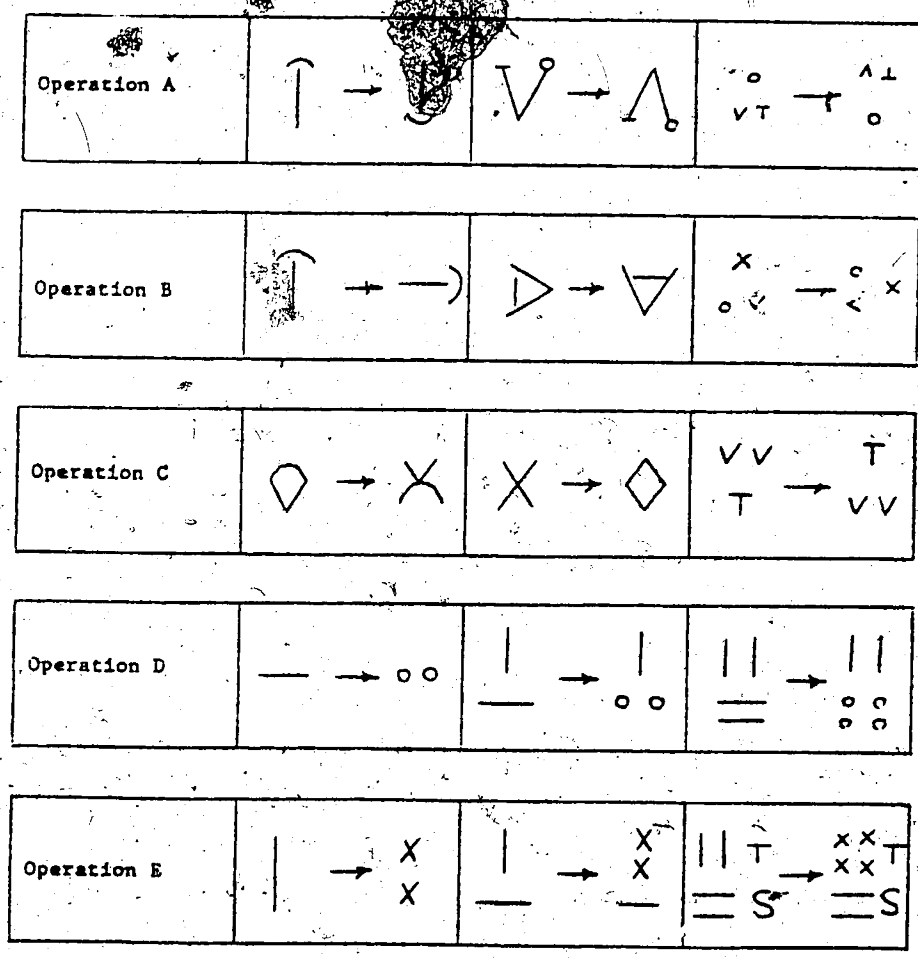
Do Operation 2 on these.



SK6: DEMONSTRATION SHEET

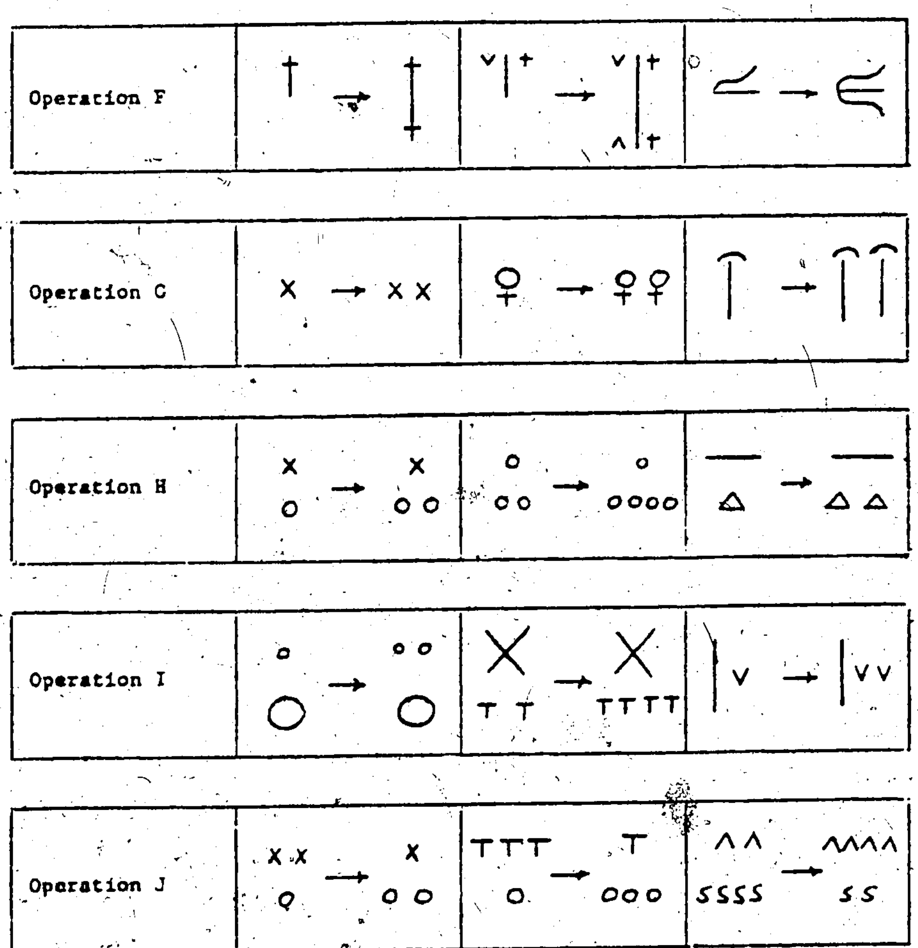
OPERATIONS A TO E

(OPERATIONS F TO J ARE ON THE NEXT PAGE)



SK6: DEMONSTRATION SHEET

OPERATIONS F TO J

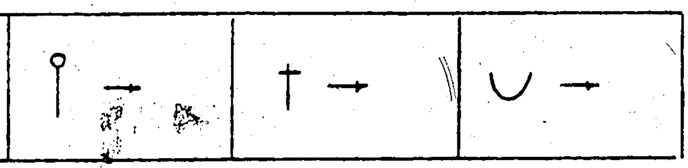
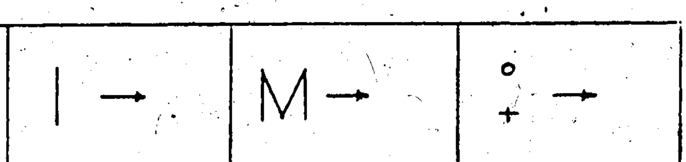
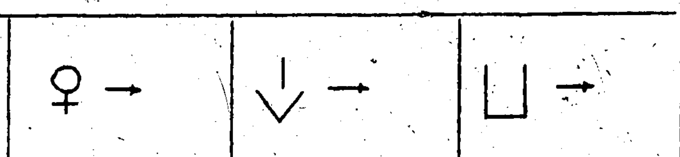
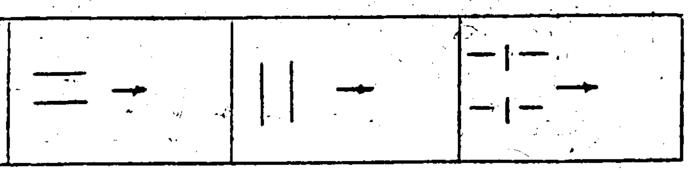
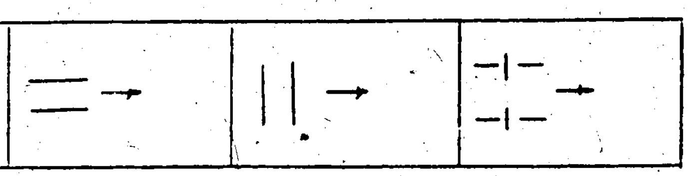


NAME .... SCHOOL ....
Last First Middle

AGE .... GRADE........... BOY GIRL DATE...............
Years (Circle one) Day Month Yr.

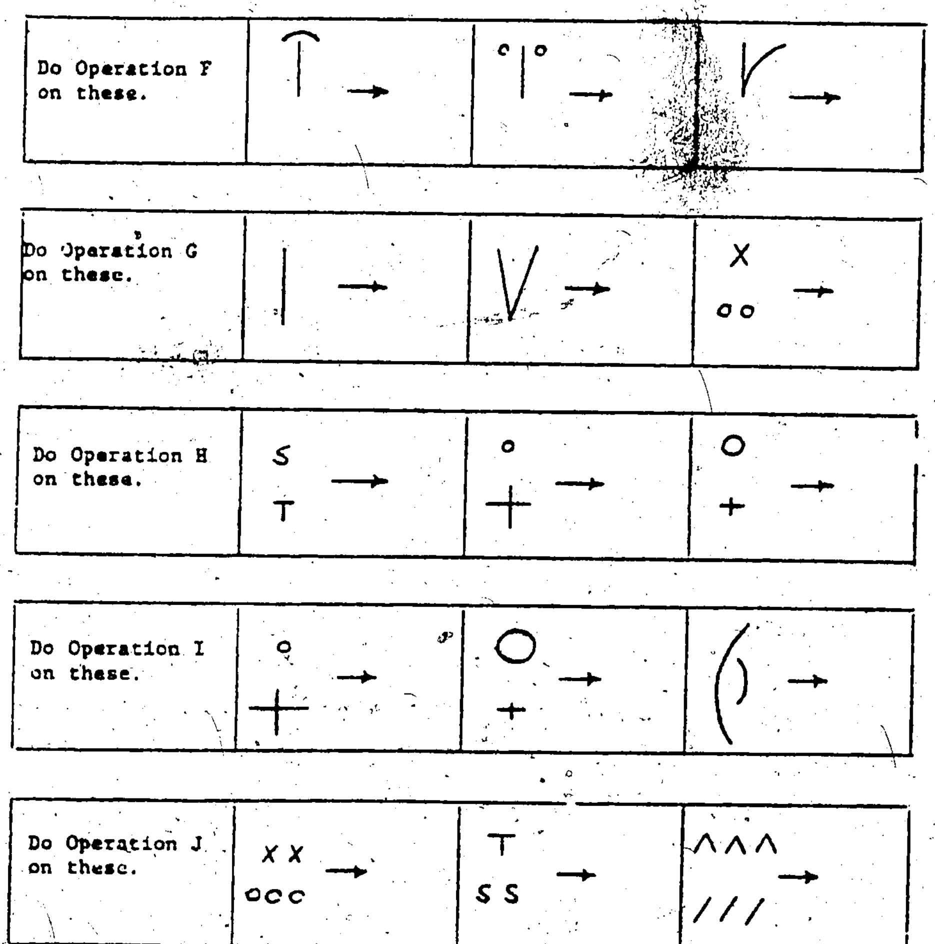
SK6: PART I

Find out the operations from the DEMONSTRATION SHEET, and fill in the answers in the blank spaces, just as you did on the PRACTICE SHEET.

Do Operation A on these.	
Do Operation B on these.	
Do Operation C on these.	
Do Operation D on these.	
Do Operation E on these.	

SKS PART I

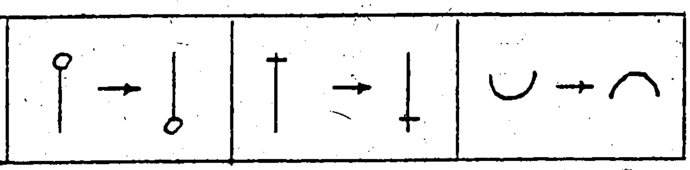
(CONTINUED)



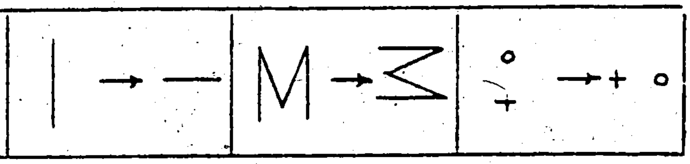
ANSWER SHEET FOR SK6, PART I

Here are the answers to the problems you did. Go through these carefully and put a tick in the right hand margin if you think that you got the whole line right. If you are not sure, ask for an explanation.

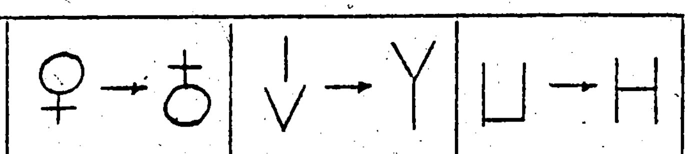
Operation A is: turn the other way up.



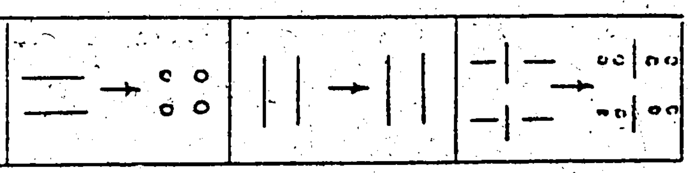
Operation B is: rotate a quarter turn clockwise.



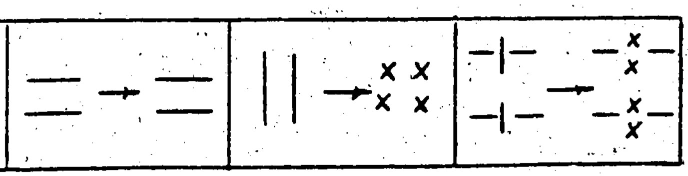
Operation C is: interchange upper and lower parts.



Operation D is: replace each horizontal line by two circles.

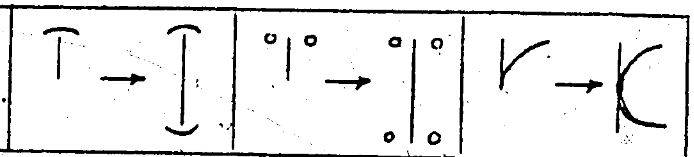
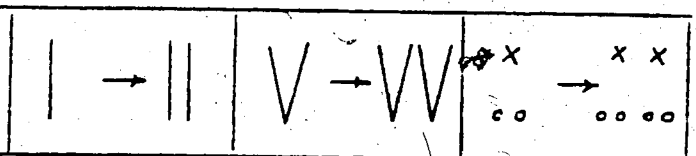
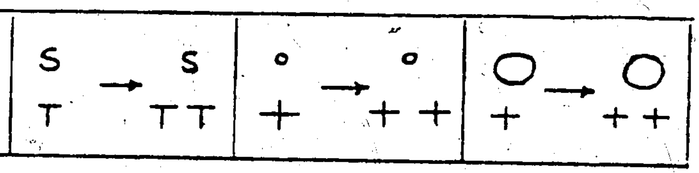
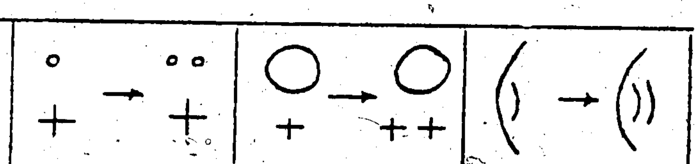
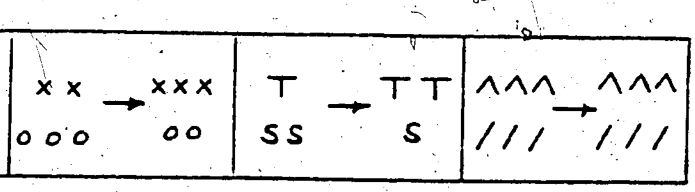


Operation E is: replace each vertical line by two crosses.



ANSWER SHEET FOR SK6, PART I

(CONTINUED)

Operation F is: add a symmetrical lower half.	
Operation G is: double everything.	
Operation H is double the lower part.	
Operation I is: double the smaller part.	
Operation J is: interchange the numbers.	

NAME.. SCHOOL.................
Last First Middle

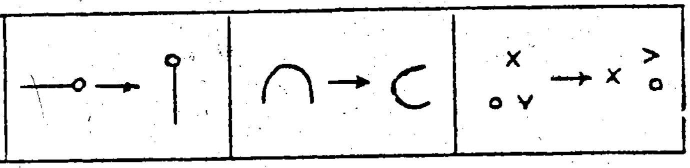
AGE............ GRADE............... BOY GIRL DATE..............
Years (Circle one) Day Month Yr.

SK6: PART II

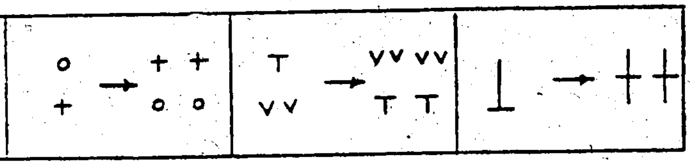
In PART II the problem is to combine the operations on the DEMONSTRATION SHEET, or to do them in reverse, or both. When combining operations, they are to be done in the order given (i.e., "Combine C and G" means "Do Operation C first and then do Operation G.")

Look at the examples given below and then carry out the operations indicated on the following three pages.

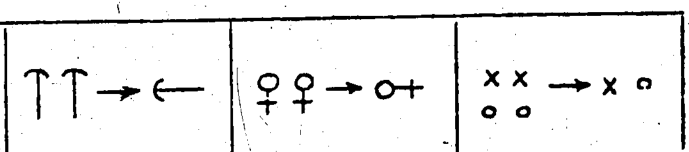
EXAMPLE:
Reverse B



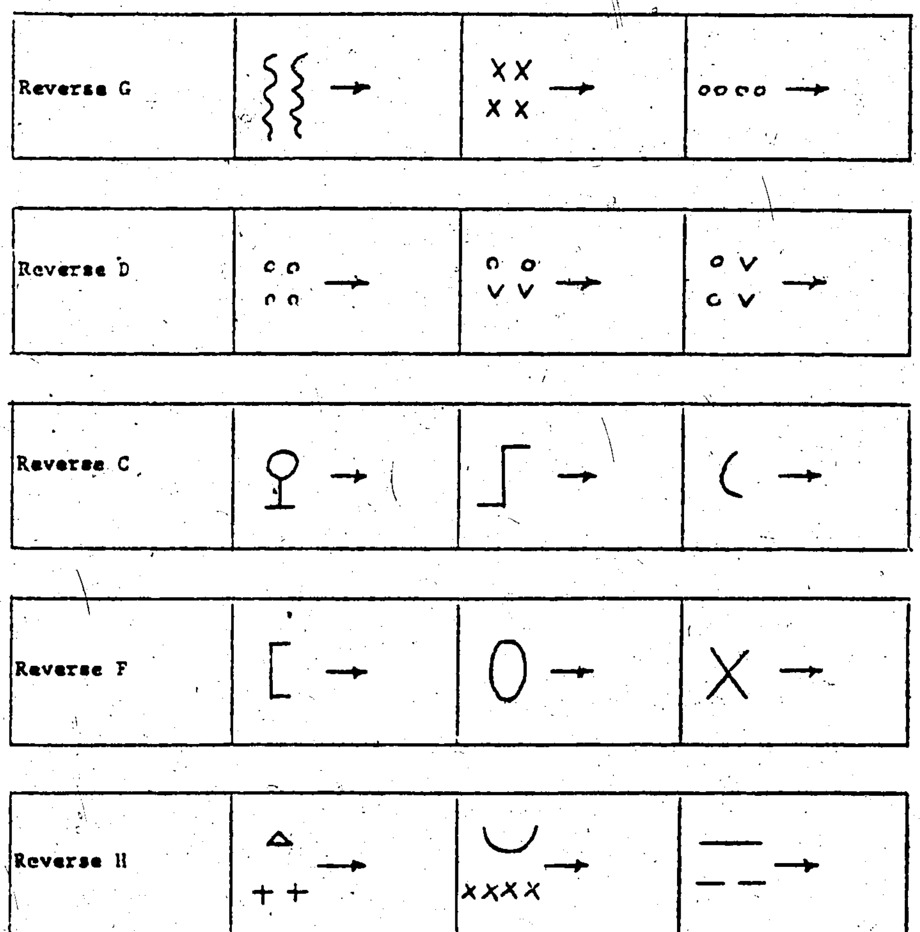
EXAMPLE:
Combine C & G



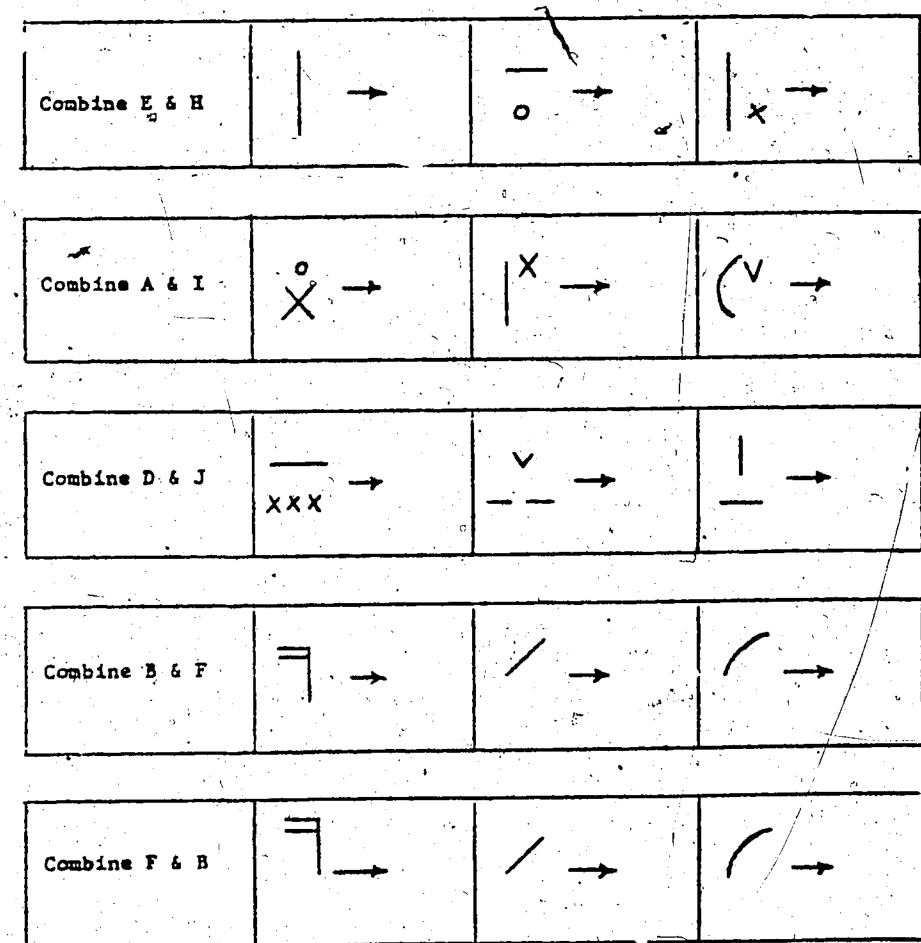
EXAMPLE:
Reverse and combine G & B



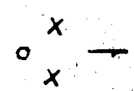
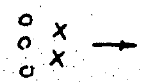
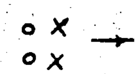
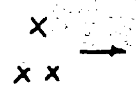
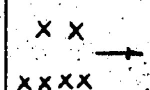
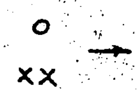
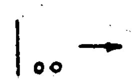
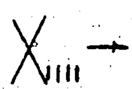
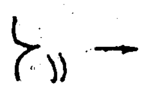
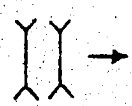
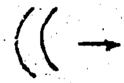
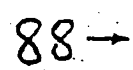
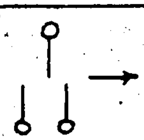
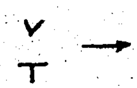
SK6: PART II



SK6: PART II



SK6: PART II

Reverse and Combine B & J			
Reverse and Combine R & E			
Reverse and Combine A & I			
Reverse and Combine F & G			
Reverse and Combine A & C			

APPENDIX 7

PROTOCOLS FOR ADMINISTRATION OF THE TEST OF DEVELOPMENTAL LEVEL AND THE SKEMP TEST

The following protocols were used for the administration of the Test of Developmental Level and the Skemp Test. The page numbers have been changed to correspond with those in Appendices 5 and 6.

Test of Developmental Level

1. "What we are going to do is not a test but a series of tasks designed to help us find out how you deal with certain kinds of situations."
2. "Please begin by putting your name on the front page of the booklet. If you have any questions now or during this period please direct them to me by raising your hand."
3. "Now turn to page 262. We will begin with the task called 'The Two Solids' because it requires demonstration by me."
4. "Question one asks you to compare the solids."

   A brass cylinder and an identical aluminum cylinder was displayed. Several members of the class in different parts of the room were asked to handle and examine the cylinders and to describe as many properties of them as they could. In each case the similarity in size and shape and the difference in weight were noted, as well as extraneous characteristics such as colour, sheen, and attachment of thread.
5. Attention was drawn to a graduated cylinder containing coloured water. Several subjects were asked to read the level of the water and, after complete agreement, the subjects were asked to write down this number as the answer to question two.

6. The aluminum cylinder (A) was carefully lowered into the water in the graduated cylinder until the aluminum cylinder rested on the bottom, completely submerged. Again several subjects were asked to determine the level of the water. This was recorded as the answer to question three.
7. Subjects were then asked to answer questions four and five.
8. When all members of the group had completed these questions the aluminum cylinder (A) was removed and the brass cylinder (B) lowered into the water. The new water level was ascertained by several subjects, and the group asked to respond to question seven if they needed to and could. At this point the class was reminded not to change any of their previous answers.
9. "Now turn to page 264." The use of the paper clips was demonstrated and the class then asked to answer the remaining three tasks (including one which was deleted from the analysis and is not shown in Appendix 5). Subjects were asked not to discuss the items and to remain in their seats until everyone had finished.

The Skemp Test

1. "Please complete the information requested on page 270. Pages 272 and 273 have been separated. Leave them face down until told to refer to them. Once again this is not a test which will count against you in any way. Do what you can. Don't worry if you are unsure of some items. The first parts are for practice and we will go on to the later part after we've talked about the early answers."
2. "Let's look at page 266. Any questions?" Questions were dealt with as they arose.

3. After all subjects were satisfied with the illustrations on page 267 they were asked to turn to page 268.
4. "On page 268 and 269 are ten more operations. What you have to do is find out what each operation is and apply it to the questions on pages 270 and 271. I suggest you find out A and apply it, find out B and apply it, etc. Do not turn to page 272 until told.
5. When all subjects had completed as much as they could of pages 268 to 271 they were asked to turn over pages 272 and 273 and to compare their answers with the correct answers indicated on these pages. Any discrepancies were discussed with the whole group.
6. The group was asked to turn to page 274 and the examples on that page were discussed until each member of the group agreed that he or she understood the examples. They were then asked to complete as many items as they could on pages 275-277, and to remain seated without talking until everyone had finished.
7. When everyone had finished the subjects were thanked for their co-operation and asked not to keep any of the pages.

APPENDIX 8

REMEDIAL UNIT 1

NAME________________________

SCHOOL______________________

Atomic Weights and Moles

Atomic Weights

The end result of your present work is the ability to calculate numbers of atoms in given weights of elements, and similarly with numbers of molecules in given weights of compounds. Chemists need to be able to do this because chemical reactions take place between atoms and molecules, and it is important for them to know just how many atoms and molecules react. You will need this skill to learn to do other things in this course, and in any future chemistry you do. A little time spent now may save you a lot later. But there is a problem . . . atoms and molecules are so small that we cannot see them, let alone count them out, as we would count apples or oranges for example. To help in this chemists use the 'mole.'

The usefulness of the 'mole' is that it allows us to count atoms and molecules indirectly. That is, we can say with reasonable certainty that a given weight of an element contains a certain number of atoms, or that a given weight of a compound contains a certain number of molecules. Moles are the link that allow us to do this.

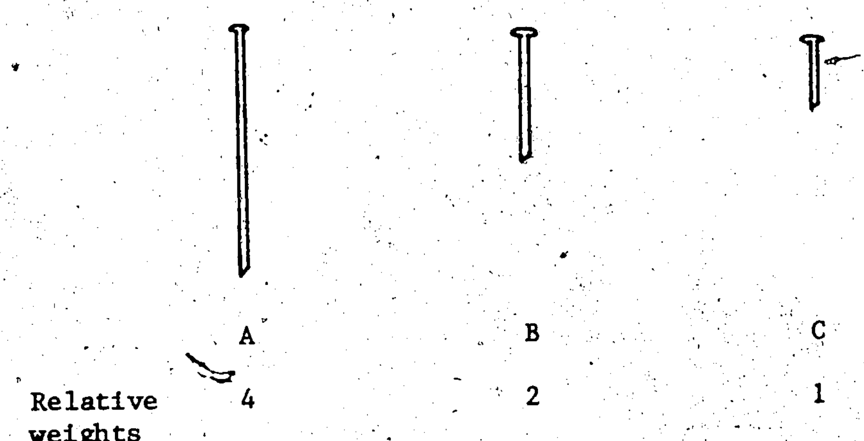
Let's begin by considering just how small atoms are. Most molecules are about the same size as atoms but just a bit bigger. To get some idea of the size of an atom consider the following approximate sizes.

a boy 100cm tall	10^2 cm
1 cm on a ruler	1 cm
a fine grain of sand	10^{-2} cm
a virus	10^{-4} cm
a large molecule	10^{-6} cm
an atom	10^{-8} cm

Note that each time we divide by one hundred. Try to divide one centimetre on a ruler one hundred times. You can just about do it. That's about the size of a fine grain of sand. Notice that to get down to the size of an atom you have to divide the grain of sand 100 x 100 x 100 times. That is one million times more. No wonder we cannot count the number of atoms in even a few grains of an element!

Although we cannot see atoms, and cannot weigh them individually, it is possible to compare their weights. Consider the following illustration.

Think of three kinds of nails. Let's call them type A, B, and C. We do not know the specific weight of any of them, but we do know that A is twice as heavy as B and that A is four times as heavy as C. That is, the relative weights are 4: 2: 1.



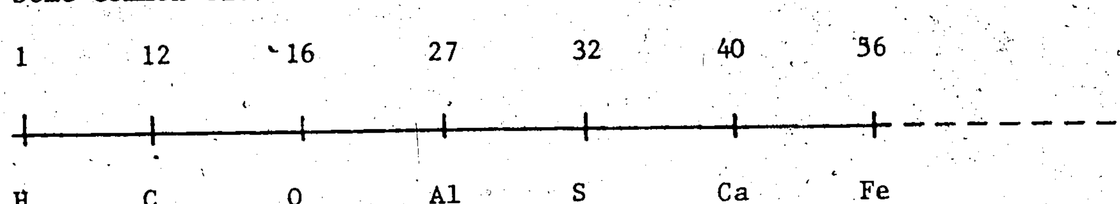
Now think of small piles of each kind of nail with 10 nails in each. How would the weights of the piles compare?

The weights are now 40 : 20 : 10

BUT the RELATIVE WEIGHTS are still 4 : 2 : 1

In general, as long as you have the SAME NUMBER of each kind, the relative weights of the piles will equal the relative weights of one nail of each kind.

The same reasoning applies to weights of atoms. The atomic weight of an element is NOT the weight of one atom of the element. It is the weight of an atom of the element relative to the weight of an atom of some standard element. The standard has changed several times in the history of chemistry, but the principle is the same. When we talk of the atomic weight of an element we are comparing the weight of one atom of the element to the weight of one atom of the standard element, or more realistically, we are comparing the weight of a large number of atoms of the element to the weight of the SAME large number of atoms of the standard. Today the standard is carbon. More correctly the weight of the carbon 12 isotope is given a value of 12.0000 and everything else is compared to it. (Do not worry about the meaning of 'isotope' . . . it will become clear later.) We therefore arrive at the atomic weight scale. If we round off the atomic weights of some common elements we find the following as part of the scale.



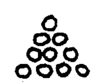
Mose chemistry books have an atomic weight table. Look up those for potassium (K), bromine (Br) and silver (Ag)

K = ____________ Br = ____________ , Ag = ____________

Moles

Once we have the idea of atomic weight established, we can begin to think of moles. Again let's begin with an illustration: Suppose you are provided with a pile of small beads. You are told that the weight of one bead is exactly one gram. You are asked to find the number of beads without counting them. What would you do?

Well, if you could weigh the beads you would be able to find the number by a simple calculation. Suppose the pile weighed say 10 grams. How many beads would that be? Well, if each weighs one gram, it would take 10 to weigh 10 grams. Therefore there are 10 beads.

	o	
weight	1 gram	10 grams
number	1	10

Let's take another example, only very slightly more difficult. Suppose a dozen apples weighs 5 lbs. How many apples would there be in a pile known to weigh 35 lbs. You would work it out this way.

5 lbs. is the weight of 1 dozen

35 lbs. is the weight of $\frac{35}{5}$ dozen

i.e., 7 dozen apples (or 7 x 12 = 84 apples).

We use the same method for using weights to count out atoms. The <u>mole</u> is the chemist's dozen, except that the number is not twelve but much larger. A mole of atoms is 6.02×10^{23} atoms. The number 6.02×10^{23} is also called the Avogadro number (N). You might be wondering what this has to do with counting out atoms. Consider the following:

It is possible for chemists to show that

12 grams of carbon contains 6.02×10^{23} atoms.

Also, 32 grams of sulfur contains 6.02×10^{23} atoms.

Do you see the connection? No? Think back to the meaning of atomic weight, and the significance of 6.02×10^{23} atoms. Let us try one more.

24 grams of magnesium contains 6.02×10^{23} atoms.

Notice what is happening.

1. We begin with the atomic weight of an element.
2. We take <u>that number</u> of grams of it.
3. It takes <u>6.02×10^{23}</u> atoms of the element to make up that weight.
4. This is <u>one mole</u> of the element.

If you weigh out the number of grams in one atomic weight, you have weighed one mole. This contains 6.02×10^{23} atoms. THIS APPLIES TO ALL ELEMENTS.

When we speak of a mole of an element we are always referring to the same number of atoms no matter what the element. We can always determine what weight of the element this is by looking up the atomic weight, and taking that number of grams. Prove to yourself you understand this by trying the following questions (self test).

1. What is the atomic weight of aluminum? ________________
2. What weight of aluminum represents one mole? ________________
3. How many atoms does this contain? ________________
4. What is the atomic weight of nitrogen? ________________
5. How many moles is this? ________________
6. How many atoms does this contain? ________________

7. Suppose the atoms of two elements were being compared. Each atom of the first element was twice as heavy as each atom of the second element. If the atomic weight of the second element is 12, what is the atomic weight of the first element? ____________

(Check your answers at the end of the booklet)

Molecular weights and moles of compounds

Now, if you have the idea of atomic weights and moles of elements, you can easily extend this to molecular weights and moles of compounds.

A compound is a number of elements joined together to produce a substance with properties different to the elements which it contains. You will find out how this happens if you study chemistry further.

Chemical reactions may be between elements, or between compounds, or between elements and compounds. Therefore, it makes sense to use the same scale for molecular weights of compounds as for atomic weights of elements; that is, the standard for each is the carbon twelve isotope taken as 12.0000 (see earlier).

Now consider what the formula of a compound tells us. It tells us the actual number of atoms that makes up one molecule of the compound.

For example, the formula of hydrogen chloride is HCl. One molecule contains one atom of hydrogen and one atom of chlorine joined together.

The formula for sulfur dioxide is SO_2. One molecule contains one sulfur atom and two oxygen atoms. In SO_2 the 2 refers only to the oxygen, i.e., to the symbol before it. Further, one MOLE of SO_2 contains one MOLE of sulfur atoms and TWO moles of oxygen atoms, while one MOLE of HCl contains one mole of hydrogen atoms and one mole of chlorine atoms. Finally, the formula of water is H_2O. One mole of H_2O contains TWO moles of hydrogen atoms and one mole of oxygen atoms. NOTE. The two in this case refers to the hydrogen. In all formulas the numbers inside or at the end of the formula always refer to the number of atoms of the element immediately in front of the number.

Now, if one mole of water contains 2 moles of hydrogen and one mole of oxygen, then the weight of one mole of water is made up of 2 x 1 grams (for the hydrogens) PLUS 16 grams for the oxygen......... that is 18 grams, or the molecular weight is 18.

The rule for calculating the molecular weight (or the number of grams in a mole) of a compound is to

1. Multiple the atomic weight of each element by the number of atoms of the element in one molecule of the compound.
2. Add the results together.

This is the molecular weight.

This number of grams is the weight of one mole.

It contains 6.02×10^{23} molecules of the compound.

Consider this example:

What is the weight of one mole of the compound phosphine, which has the formula PH_3?

1. the atomic weight of phosphorus is 31
2. the atomic weight of hydrogen is 1, multiplied by 3 gives 3
3. the total is 31 + 3 = 34
4. the molecular weight is 34
5. the weight of one mole is 34 grams
6. 34 grams of phosphine contains 6.02×10^{23} molecules.

Now try self-test (2):

1. What is molecular weight of hydrogen bromide (HBr)? __________
2. What is the weight of one mole of it? __________
3. How many molecules does it contain? __________
4. What is the molecular weight of carbon disulfide (CS_2)? ______
5. What is the weight of one mole? __________
6. How many molecules does it contain? __________

Answers to self-test (1):

1. 27 2. 27 grams 3. 6.02×10^{23} 4. 14
5. one 6. 6.02×10^{23} 7. 24 (2x12).

Answers to self-test (2):

1. 81 (1+80) 2. 81 grams 3. 6.02×10^{23}
4. 76 (12+32+32) 5. 76 grams 6. 6.02×10^{23}

APPENDIX 9

REMEDIAL UNIT 2

NAME____________________

SCHOOL__________________

Masses, Moles, and Atoms

If you are reading this, you must be having some trouble either with the link between masses and moles or between moles and atoms, or both. It is important for you to understand this as it will affect your learning of other chemistry concepts. Hopefully you have some understanding of the meaning of the term 'mole' from your previous work. All that you have to do now is to apply this understanding to some specific examples. So let's try again.

Moles ⟷ Atoms or Molecules

Let's begin by recalling the meaning of the term mole. Remember that the mole is in a sense the chemist's dozen except that the number represented by one mole is 6.02×10^{23} atoms of an element or 6.02×10^{23} molecules of a compound. Let's illustrate with 'dozen' and then apply to 'mole.'

Consider the question........How many eggs are there in 3 dozen?

You would reason:

one dozen eggs is twelve eggs,

Therefore, 3 dozen eggs is 3 x 12 eggs, which is 36 eggs.

Similarly half a dozen eggs is ½ x 12 eggs, which is 6 eggs.

Now consider this question.........How many atoms are there in 3 moles of an element? This can be answered just as above:

one mole is 6.02×10^{23} atoms.

Therefore 3 moles is $3 \times 6.02 \times 10^{23}$ atoms,

which is 18.06×10^{23} atoms, or 1.806×10^{24} atoms.

Similarly, half a mole would contain $0.5 \times 6.02 \times 10^{23}$ atoms, which is 3.01×10^{23} atoms.

Now consider the opposite kind of calculation, changing number of atoms to number of moles.

Consider the question.........How many dozen apples are there in 60 apples?

You would reason:

12 apples makes 1 dozen.

Therefore 1 apple makes 1/12 dozen,

and 60 apples makes 60/12 dozen,

which is 5 dozen.

Now consider the question.........How many moles are there in 6.02×10^{24} atoms of an element?

You would reason:

6.02×10^{23} atoms makes 1 mole.

1 atom makes $\frac{1}{6.02 \times 10^{23}}$ moles.

6.02×10^{24} atoms makes $\frac{6.02 \times 10^{24}}{6.02 \times 10^{23}}$ moles

which is 10 moles.

All of the above applies equally well to moles of compounds and the number of molecules contained, because ONE MOLE of a compound contains 6.02×10^{23} molecules.

For example,

1. How many molecules are there in 3 moles of carbon dioxide?

   1 mole contains 6.02×10^{23} molecules. Therefore:

   3 moles contains $3 \times 6.02 \times 10^{23}$ molecules,

   which is 18.06×10^{23}, or 1.806×10^{24} molecules

2. How many moles are there in 3.01×10^{23} molecules of carbon dioxide?

   6.02×10^{23} molecules is 1 mole,

   $$1 \text{ molecule is } \frac{1}{6.02 \times 10^{23}} \text{ moles,}$$

   $$3.01 \times 10^{23} \text{ molecules is } \frac{3.01 \times 10^{23}}{6.02 \times 10^{23}} \text{ moles,}$$

   which is 0.5 moles.

Now for a simple extension:

One mole of any element contains 6.02×10^{23} atoms.

e.g., 1 mole of carbon contains 6.02×10^{23} atoms,

1 mole of iron contains 6.02×10^{23} atoms.

We can generalize this as follows:

ONE MOLE of any element contains THE SAME number of atoms as one mole of any element (we know this to be 6.02×10^{23} atoms, but let's not worry about this specific number for the moment).

Now, suppose we take half a mole of several different elements. They no longer contain 6.02×10^{23} atoms but they all contain the SAME number of atoms as each other.

In the same way, if we had, say, 0.157 moles of different elements, each contains the same number of atoms. (No longer 6.02×10^{23}, and we are not really interested in exactly how many.)

Let's generalize

SAME NUMBER OF MOLES means

SAME NUMBER OF ATOMS

............and

SAME NUMBER OF ATOMS means

SAME NUMBER OF MOLES

The same is true for compounds except that we refer to molecules instead of atoms.

Thus for compounds, EQUAL NUMBERS OF MOLES means EQUAL NUMBERS OF MOLECULES, and vice-versa.

To take this one small step further,

Twice as many moles would mean twice as many atoms of an element or twice as many molecules of a compound; one-third as many moles would mean one-third as many atoms or molecules; etc.

For example:

Two moles of copper contains exactly as many atoms as two moles of iron, and twice as many atoms as one mole of sulfur.

One third of a mole of carbon dioxide contains exactly the same number of molecules as one-third of a mole of methane (CH_4).

Masses ⟷ Moles

Hopefully you can now convert moles of elements and compounds to numbers of atoms and molecules. However, when we weigh out substances we deal in grams not moles. How can we convert grams to moles, and also if we want a particular number of moles how can we calculate how many grams we need? This is the focus of this section.

Recall how you earlier determined the mass of one mole of an element. All you had to do was find the atomic weight and weigh out that many grams.

For compounds it was a little more difficult.

The procedure you followed was:

1. Multiply the atomic weight of each element present by the number of atoms of the element in one molecule of the compound (remembering that any number inside the formula or at the end of it refers only to the number of atoms immediately in front of the number).
2. Add the results together.
3. This many grams is the weight of one mole.

For example, the molecular weight of propane (C_3H_8) is 44. You could calculate it as follows:

1. The atomic weight of carbon is 12. The formula for propane tells us that one molecule contains 3 carbon atoms. The contribution of carbon to the molecular weight is therefore 3 x 12, or 36 atomic weight units.
2. The atomic weight of hydrogen is 1.

   The 8 hydrogen atoms in the molecule therefore contribute 8 x 1, or 8 atomic weight units.

3. The total is 36 x 8 = 44 units on the atomic weight scale. Therefore, the molecular weight is 44, and the weight of one mole is 44 grams.

Now, once you can do this, you should be able to calculate the mass of any number of moles.

For example, the mass of 5 moles of carbon is <u>five</u> times the mass of <u>one mole</u>, i.e., 5 x 12 = 60 grams.

The mass of one-half a mole of propane is $\frac{1}{2}$ x 44 = 22 grams.

The rule is very simple.

TO CONVERT MOLES TO MASS, <u>MULTIPLY</u> THE NUMBER OF MOLES BY THE MASS OF ONE MOLE.

The opposite is equally simple.

TO CONVERT MASS TO MOLES, <u>DIVIDE</u> THE MASS (NUMBER OF GRAMS) BY THE MASS OF ONE MOLE.

e.g., How many moles are there in 48 grams of carbon?

12 grams of carbon is 1 mole

48 grams of carbon is 48/12 moles, or 4 moles

If you understood the content of this booklet, you can now:

1. Convert moles of elements and compounds to the equivalent numbers of grams.
2. Convert grams of elements and compounds to the equivalent number of moles.
3. Relate different numbers of moles of substances to the equivalent numbers of atoms or molecules present.
4. Convert numbers of moles of substances to numbers of atoms or molecules present in elements and compounds respectively.

5. Relate different numbers of moles to the equivalent numbers of atoms or molecules.

To see if you can do these things try the following self-test. If you are unsuccessful on any questions go through this booklet again. It will really pay off later.

<u>Self-Test</u>. (Cover the answers until you have finished)

1. How many moles are represented by:
   (a) 78 grams potassium (K)?
   (b) 17 grams hydrogen sulfide (H_2S)?
2. What is the weight of:
   (a) 3 moles carbon (C)?
   (b) half a mole of ammonia (NH_3)?
3. How many atoms are there in 3 moles of sulfur (S)?
4. How many molecules are there in half a mole of methane (CH_4)?
5. How many moles of the named substance are there in:
   (a) 6.02×10^{27} atoms of magnesium (Mg)
   (b) 3.01×10^{23} molecules of carbon monoxide (CO)
6. Which of the following contains most atoms:
   (a) 24g carbon (C)   (b) 27g aluminum (Al)
   (c) 48g sulfur (S)

<u>Answers</u>

1. 2, 0.5
2. 36g, 8.5g
3. 1.806×10^{24} (i.e. $3 \times 6.02 \times 10^{23}$)
4. 3.01×10^{23}
5. 10^4, 0.5
6. 24gC (2 moles)

APPENDIX 10

TABLE 37

Test Scores for Skills A to H on the Final Test and Quizzes, the Test of Developmental Level (DL) and the Skemp Test (SK)

NOTE: Subjects whose responses were not clearly interpretable for a sub-test of the Final Test or a quiz were treated as missing data for analyses involving the particular sub-test(s). Such subjects are represented by a '9' in Table 37. Subjects who were absent for any quiz were eliminated from the sample. Subjects who were absent from any other test are represented by a '-' in Table 37 for the particular test(s). Subjects whose ID begins with 1 or 4 were in the remedial groups.

ID	Final Test A	B	C	D	E	F	G	H	Quizzes A	B	C	D	E	F	G	H	DL TS	RT	RA	DL	SK
101	8	6	8	8	8	8	8	4	8	6	8	5	6	8	8	4	2	2	2	6	29
102	8	8	8	8	8	8	8	4	8	8	4	8	8	8	2	4	2	2	2	6	22
103	2	2	4	4	2	7	2	4	0	2	8	5	6	8	0	4	2	2	0	4	14
104	8	8	6	8	8	8	8	4	9	9	2	8	4	8	2	4	0	1	2	3	21
105	8	8	8	8	8	8	8	4	8	6	4	8	4	8	6	4	2	2	0	4	22
106	8	4	8	8	6	8	8	4	8	4	8	7	4	6	8	4	1	2	1	4	21
107	8	8	4	7	8	7	8	4	8	4	9	8	9	8	2	4	0	2	2	4	32
108	0	6	0	6	6	8	6	4	9	9	2	3	0	8	0	4	-	-	-	-	--
109	6	6	8	8	8	8	8	4	8	8	4	8	6	8	0	4	2	1	2	5	26
110	8	8	8	8	8	8	8	4	8	6	8	8	8	8	8	4	-	-	-	-	--
111	8	6	4	6	8	8	8	4	8	8	8	8	8	8	0	4	1	2	0	3	23
112	8	8	8	8	8	8	8	4	8	8	8	8	8	8	0	4	2	1	0	3	24
113	8	8	8	8	8	8	8	4	8	8	8	8	8	8	8	4	1	1	1	3	18
114	8	8	8	8	8	8	8	4	0	8	4	8	8	7	8	4	1	2	2	5	23
115	8	6	6	6	6	8	8	4	4	4	8	8	2	8	2	4	2	2	2	6	11
116	8	2	4	3	8	8	4	4	0	0	0	5	2	8	0	4	2	2	2	6	24
117	4	6	6	8	6	8	8	4	0	0	0	0	0	0	0	4	1	0	0	1	9
118	8	2	2	4	8	7	8	4	6	4	0	3	0	0	0	4	0	0	1	1	10
119	4	4	0	2	8	6	0	4	2	2	9	1	9	9	2	4	1	1	0	2	15
120	6	8	8	6	6	8	8	2	8	6	6	8	2	7	6	4	2	1	2	5	0
121	8	8	8	8	8	8	8	4	8	6	4	7	8	6	0	4	2	2	1	5	18
122	8	8	8	8	8	8	8	4	0	8	9	0	9	6	0	4	1	2	2	5	19
123	8	8	8	8	8	8	8	4	8	8	8	8	8	8	8	4	2	1	2	5	27
124	8	8	8	8	8	7	8	4	8	8	8	8	8	8	0	4	2	2	2	6	35
125	0	6	2	4	6	7	6	4	0	0	8	6	6	8	0	4	1	1	0	2	17
126	8	8	8	8	8	8	2	4	6	4	8	8	8	8	4	4	1	2	0	3	22
127	8	8	8	8	6	8	8	4	8	8	8	8	8	8	2	4	2	1	2	5	21
128	8	6	8	6	6	8	8	4	0	2	0	6	0	8	8	4	2	1	0	3	2
129	6	8	6	8	8	8	8	4	8	6	2	8	8	8	0	4	0	1	2	3	17
130	8	8	8	8	8	8	8	4	8	6	8	6	8	6	8	4	2	1	2	5	19
131	0	6	4	4	0	6	2	4	0	0	2	1	0	2	0	2	0	1	0	1	1
132	8	8	6	8	8	8	8	4	8	8	8	8	6	8	0	4	2	1	2	5	28
133	0	6	2	0	2	5	4	4	6	6	8	7	8	8	2	4	1	1	0	2	1
134	8	8	8	8	8	8	8	4	8	6	8	8	6	8	4	4	2	2	2	6	12
135	4	2	8	6	8	8	6	4	9	9	8	6	8	8	0	4	2	1	1	4	--

....continued

ID	Final Test A	Final Test B	Final Test C	Final Test D	Final Test E	Final Test F	Final Test G	Final Test H	Quizzes A	Quizzes B	Quizzes C	Quizzes D	Quizzes E	Quizzes F	Quizzes G	Quizzes H	DL TS	DL RT	DL RA	DL DL	SK
136	6	8	8	8	8	8	8	4	8	6	2	8	8	8	8	4	2	1	1	4	25
137	8	8	8	8	8	8	8	4	0	0	8	8	6	8	8	4	2	2	2	6	19
138	8	8	4	8	8	7	4	4	8	2	2	8	6	8	2	4	1	2	2	5	15
139	6	6	8	8	8	7	8	2	9	9	0	0	0	7	2	4	2	2	2	6	27
140	6	4	4	5	6	8	8	8	4	2	2	6	4	8	0	4	2	0	1	3	--
141	2	8	6	8	8	8	6	4	0	0	2	8	8	8	4	4	2	2	2	6	18
201	6	4	6	8	8	8	6	4	0	0	0	0	0	2	4	4	1	1	2	4	15
202	0	0	0	0	0	2	4	4	0	0	9	0	9	0	0	2	1	0	0	1	0
203	6	8	0	6	8	8	0	0	0	6	0	2	4	8	0	0	2	2	2	6	16
204	0	0	0	0	0	6	4	4	0	0	0	0	0	8	4	2	1	2	2	5	21
205	0	0	0	0	0	2	0	0	0	0	2	2	4	0	0	4	2	1	1	4	17
206	8	2	8	6	8	8	8	4	8	4	4	6	8	8	0	4	2	1	2	5	24
207	8	8	8	8	6	8	8	2	8	8	8	8	4	8	4	4	-	-	-	-	--
208	8	8	8	8	8	8	8	4	6	4	8	8	8	8	8	4	2	1	2	5	36
209	8	8	8	8	8	8	8	4	8	8	8	8	8	8	8	4	2	2	2	6	32
210	6	8	4	8	8	8	8	4	8	8	8	3	8	8	8	4	2	2	2	6	27
211	6	2	0	0	8	8	0	4	0	0	9	9	9	9	0	4	0	2	2	4	8
212	4	6	0	4	2	8	2	2	2	4	0	4	2	8	0	4	2	1	2	5	12
213	2	0	0	0	8	8	4	4	0	2	0	0	8	8	0	4	0	2	2	4	19
214	6	4	0	4	8	8	2	4	0	0	0	5	8	8	0	4	2	1	2	5	24
215	6	6	8	6	8	8	8	4	8	2	8	8	6	8	8	4	2	2	2	6	29
216	0	0	2	0	4	6	0	2	4	0	0	0	0	8	0	4	0	2	2	4	11
217	2	0	0	1	0	8	6	4	0	0	0	2	4	8	4	4	2	2	2	6	18
218	0	6	0	2	8	7	6	2	0	0	0	5	8	7	0	2	2	2	1	5	15
219	2	2	2	2	0	8	4	2	0	0	2	0	0	6	0	4	1	0	0	1	7
220	6	8	6	8	8	8	8	4	4	6	2	8	8	8	4	4	2	1	2	5	38
221	0	2	0	0	6	8	0	2	2	6	0	2	8	8	4	0	2	2	2	6	23
222	8	8	0	8	8	8	8	4	8	4	4	6	8	8	0	2	2	1	1	4	27
223	8	8	4	8	8	8	8	4	4	4	4	4	8	8	8	4	1	1	1	3	17
224	0	8	0	8	8	8	4	4	0	4	2	6	6	8	0	4	2	0	0	2	9
225	6	8	8	8	6	8	8	4	9	9	8	8	8	8	8	4	2	2	2	6	24
226	2	0	0	0	4	3	2	4	6	0	9	4	4	5	2	2	1	1	2	4	3
227	6	6	6	8	6	8	8	4	8	8	8	4	6	7	8	4	2	1	2	5	26
250	9	6	0	4	9	9	9	4	8	8	9	0	9	0	2	4	2	2	2	6	24
251	-	-	-	-	-	-	-	-	8	0	9	2	9	7	8	4	2	1	1	4	0

....continued

	Final Test								Quizzes								DL				
ID	A	B	C	D	E	F	G	H	A	B	C	D	E	F	G	H	TS	RT	RA	DL	SK
252	–	–	–	–	–	–	–	–	0	8	8	9	9	9	0	4	0	1	2	3	8
253	4	8	4	8	4	7	6	4	8	8	8	8	9	7	4	4	2	2	2	6	15
254	6	8	8	8	8	8	8	4	8	4	8	8	8	8	8	2	2	1	2	5	21
255	9	4	2	5	9	9	9	4	2	4	9	0	2	6	4	4	0	2	1	3	18
256	9	9	9	9	9	9	4	2	0	0	0	2	0	0	0	4	0	1	0	1	9
257	0	2	0	5	2	4	4	4	0	0	0	0	0	8	2	4	0	2	0	2	4
258	0	0	0	0	2	6	4	4	0	0	0	0	2	6	4	2	2	0	0	2	0
259	0	0	0	4	8	6	0	4	2	2	0	6	8	8	6	4	2	2	2	6	19
260	9	9	9	8	9	9	8	4	9	6	8	8	6	8	6	4	2	2	2	6	0
261	9	9	9	8	9	8	8	2	8	4	9	2	9	8	4	4	2	2	2	6	19
262	2	8	2	4	4	8	2	4	8	2	9	4	9	8	8	0	0	2	1	3	26
263	0	0	0	0	4	7	4	4	2	0	0	0	9	7	2	2	1	1	2	4	10
264	6	2	2	2	4	7	2	4	4	0	0	0	9	8	4	4	2	1	2	5	20
265	9	9	9	3	9	8	6	9	0	0	0	0	0	6	0	4	2	2	2	6	17
266	2	4	2	5	2	8	4	2	4	2	2	0	0	0	4	4	2	2	1	5	--
267	9	0	9	0	9	9	2	2	0	0	0	2	0	0	0	4	0	1	0	1	16
268	9	9	9	9	9	9	9	2	0	0	0	2	0	0	0	4	–	–	–	–	--
269	9	9	9	4	9	9	9	2	9	9	0	2	0	8	4	4	1	0	2	3	0
270	9	9	9	8	9	8	8	4	8	8	8	6	2	7	8	4	2	1	1	4	21
271	–	–	–	–	–	–	–	–	0	2	9	9	9	9	4	4	2	2	0	4	16
272	0	2	0	4	0	8	6	4	0	2	0	4	0	0	4	4	2	2	1	5	22
301	0	0	0	0	4	7	4	4	0	0	0	0	4	8	2	4	2	2	1	5	9
302	0	0	0	1	8	8	4	4	2	0	0	0	8	6	4	2	1	1	2	4	29
303	2	0	0	0	6	8	2	4	0	0	0	0	4	5	0	4	2	2	2	6	24
304	–	–	–	–	–	–	–	–	0	0	9	0	9	9	9	9	1	1	0	2	0
305	8	6	6	8	8	8	8	4	9	9	9	9	9	9	9	9	2	2	2	6	23
306	4	4	4	3	8	8	4	2	9	9	0	0	8	8	0	4	1	2	2	5	28
307	2	8	0	6	8	8	6	4	0	2	9	0	8	8	4	4	1	2	1	4	16
308	0	2	2	0	6	8	4	4	9	9	9	0	9	8	8	4	1	2	0	3	23
309	0	0	0	2	6	8	4	4	2	0	9	0	9	8	0	2	1	2	2	5	13
310	0	2	2	2	8	8	6	4	2	0	0	0	8	7	0	4	–	–	–	–	--
311	0	4	0	0	6	3	0	4	0	0	0	0	9	6	4	4	0	2	1	3	11
312	0	2	0	4	4	6	6	4	2	0	4	1	9	3	4	4	2	2	2	6	15
313	0	6	0	8	4	8	6	2	9	9	9	4	9	8	4	4	–	–	–	–	--

D	Final Test A	B	C	D	E	F	G	H	Quizzes A	B	C	D	E	F	G	H	DL TS	RT	RA	DL	SK
14	6	0	0	0	4	8	0	2	0	0	0	0	0	5	0	4	–	–	–	–	--
15	8	4	8	6	8	7	4	4	9	9	9	9	9	9	9	9	2	2	2	6	32
16	0	0	0	0	4	6	2	2	0	0	0	0	0	0	2	4	0	0	0	0	5
17	–	–	–	–	–	–	–	–	0	0	0	0	2	8	0	4	2	1	1	4	32
18	2	2	0	1	8	8	2	4	0	2	2	0	2	7	4	4	2	2	1	5	--
19	8	2	2	3	8	8	6	4	6	4	0	0	9	8	0	2	2	2	2	6	--
20	2	2	0	0	8	8	8	4	0	0	0	0	6	8	8	4	1	1	1	3	12
21	0	0	0	4	8	8	0	4	4	0	0	0	6	8	0	2	2	1	2	5	26
22	0	2	8	3	6	8	4	4	6	0	6	0	8	8	2	4	2	2	0	4	15
23	0	0	0	0	4	5	6	0	8	0	0	0	0	8	0	2	1	2	2	5	7
24	2	0	0	0	6	8	4	4	0	0	0	0	0	6	4	4	1	1	2	4	0
25	0	0	0	0	4	6	0	4	0	0	0	0	6	8	2	4	1	2	0	3	17
26	2	0	0	0	8	8	4	4	0	0	0	0	2	8	2	2	2	2	0	4	9
27	6	2	0	0	6	8	4	4	0	0	4	0	0	8	0	4	2	2	2	6	27
28	2	2	0	0	6	8	4	4	0	2	2	4	0	7	4	4	2	2	2	6	14
29	8	8	6	7	8	8	8	4	9	9	0	8	8	8	8	2	2	2	2	6	28
50	0	2	0	4	4	8	4	2	0	0	2	0	8	4	0	4	2	2	2	6	29
51	2	0	0	0	4	8	2	4	0	0	0	0	9	8	2	4	0	1	2	3	14
52	0	2	0	0	4	4	4	4	0	0	2	0	9	8	0	4	1	2	0	3	20
53	6	6	2	1	6	8	4	4	4	2	2	0	6	8	4	4	2	2	0	4	14
54	0	4	0	6	6	3	6	2	0	0	2	0	8	8	4	4	2	1	2	5	30
55	2	6	4	4	8	8	4	4	0	0	2	0	6	7	4	4	2	2	2	6	15
56	6	0	2	0	8	7	4	4	9	9	2	0	9	8	0	4	2	2	2	6	28
57	0	2	0	2	2	8	4	0	9	9	9	0	9	8	2	4	2	2	2	6	18
58	9	4	4	9	6	9	9	4	4	2	0	1	4	8	0	4	2	2	2	6	25
59	0	0	2	0	6	5	0	4	9	9	9	0	9	5	0	4	–	–	–	–	25
60	8	6	8	8	8	8	8	4	8	8	8	8	8	8	8	4	2	2	2	6	31
61	9	6	2	9	8	9	9	4	9	9	2	2	8	8	6	4	2	2	2	6	28
62	0	2	0	2	6	6	4	4	6	6	9	8	6	8	0	2	–	–	–	–	--
63	9	9	0	9	4	9	9	4	9	9	0	0	0	0	4	0	2	2	1	5	13
64	2	0	0	0	6	8	2	4	0	0	2	0	2	7	0	4	–	–	–	–	28
65	2	0	0	2	8	6	4	2	8	0	2	0	4	8	4	4	1	1	2	4	22
66	4	2	0	4	8	8	6	4	0	2	2	0	6	8	4	4	2	2	2	6	24
67	4	0	0	0	8	7	4	0	9	9	0	0	2	6	0	2	2	2	0	4	22

....continued

ID	Final Test A	B	C	D	E	F	G	H	Quizzes A	B	C	D	E	F	G	H	DL. TS	RT	RA	DL	SK
368	8	8	8	6	8	8	8	4	9	6	2	0	8	8	8	4	2	2	2	6	23
369	9	8	8	9	8	9	9	4	8	8	8	8	8	8	8	4	2	2	2	6	--
370	4	0	0	0	6	6	4	4	2	2	0	0	2	8	0	4	–	–	–	–	--
371	0	0	0	0	8	8	4	4	0	2	9	0	9	8	2	4	2	1	2	5	17
372	2	0	0	0	6	8	0	4	2	0	0	0	2	6	0	4	1	0	0	1	15
373	9	0	2	9	2	9	9	4	0	0	9	9	9	9	0	4	–	–	–	–	5
374	9	6	2	9	8	9	9	0	0	4	0	0	0	8	2	4	2	2	2	6	28
375	6	0	4	0	8	8	2	4	2	0	0	0	0	2	4	4	2	0	2	4	17
376	0	0	0	0	4	8	6	4	0	0	2	0	0	8	2	4	1	0	0	1	12
377	–	–	–	–	–	–	–	–	9	9	2	0	9	8	0	4	2	2	2	6	30
378	9	8	6	9	8	9	9	4	9	6	4	2	9	8	9	9	2	1	2	5	34
379	0	0	0	1	4	8	4	2	0	2	0	0	6	8	4	4	1	1	2	4	21
380	9	9	2	9	4	9	9	4	0	0	0	0	0	5	9	9	2	1	0	3	9
381	6	2	2	4	8	8	6	4	9	2	0	0	6	8	4	4	1	2	2	5	11
382	4	6	0	4	8	8	6	4	2	2	9	0	9	8	4	4	2	2	2	6	26
383	0	4	0	0	6	8	6	4	9	0	9	0	9	6	0	4	2	2	2	6	18
384	–	–	–	–	–	–	–	–	0	2	0	0	0	4	0	4	1	1	2	4	24
401	4	2	2	3	8	8	6	4	4	4	0	0	6	8	9	4	2	1	2	5	1
402	6	4	2	6	8	7	8	4	4	0	0	2	9	5	2	4	1	2	0	3	16
403	4	0	0	1	8	8	0	4	9	9	2	2	6	6	2	4	2	2	2	6	23
404	4	0	0	4	4	3	4	4	2	0	0	1	0	3	8	4	2	2	2	6	7
405	4	2	0	5	4	7	2	4	8	2	2	4	6	8	4	4	2	2	2	6	23
406	6	2	0	6	4	7	8	4	6	2	2	4	2	8	8	4	1	1	2	4	12
407	0	0	0	0	2	4	6	2	0	0	0	0	0	6	6	4	2	2	1	5	22
408	2	0	0	5	6	7	0	4	2	2	0	1	0	7	2	4	0	2	0	2	16
409	4	2	2	4	8	6	8	4	6	0	9	4	9	6	4	4	2	2	2	6	38
410	4	6	4	4	8	8	0	4	8	2	0	1	8	8	2	4	–	–	–	–	--
411	2	6	0	6	2	8	8	4	9	4	2	0	0	8	6	4	0	2	2	4	21
412	2	0	2	4	6	7	8	4	9	9	0	4	4	3	6	4	1	1	2	4	20
413	6	4	2	5	8	8	8	4	8	2	0	2	2	6	4	4	1	2	0	3	18
414	2	0	0	0	6	5	4	4	6	4	0	3	9	3	2	4	0	1	0	1	1
415	6	2	6	3	8	8	8	4	2	0	0	3	6	6	8	4	–	–	–	–	--
416	2	0	2	5	0	3	4	4	0	0	0	2	0	1	4	4	1	2	0	3	8
417	0	0	0	1	6	7	8	4	0	2	0	0	4	7	6	4	1	1	1	3	0

....continued

D	Final Test A	B	C	D	E	F	G	H	Quizzes A	B	C	D	E	F	G	H	DL TS	RT	RA	DL	SK
18	2	0	0	4	6	8	0	4	0	0	0	3	0	2	4	4	2	2	0	4	9
19	8	2	2	5	6	8	6	4	6	6	0	8	6	7	8	4	0	1	2	3	19
20	4	4	0	2	6	–	8	4	2	2	9	2	8	9	8	4	2	2	2	6	21
21	4	0	0	4	6	6	4	4	0	0	0	0	0	4	0	4	0	2	2	4	7
22	0	0	0	0	4	4	0	4	2	0	9	0	9	6	8	4	0	2	0	2	15
23	4	4	0	1	8	8	6	2	2	2	0	2	2	5	0	4	2	2	0	4	17
24	0	4	0	3	6	8	4	2	0	4	2	3	4	7	2	0	2	2	2	6	7
25	6	6	2	4	6	8	6	4	9	9	4	4	6	7	4	4	1	2	0	3	27
26	2	6	0	4	8	8	2	4	4	6	0	0	2	5	4	4	2	2	2	6	16
27	4	6	0	3	6	8	4	4	9	9	9	9	9	9	9	9	2	2	0	4	21
28	6	4	4	6	6	8	8	2	0	6	0	0	6	6	8	4	2	2	2	6	--
29	6	4	2	6	6	8	8	2	8	4	9	6	8	8	6	4	2	2	2	6	17
50	4	4	6	8	2	8	6	4	4	4	0	5	2	7	6	2	1	1	2	4	23
51	6	4	0	8	8	8	0	4	8	6	9	4	2	8	0	4	2	0	0	2	23
52	8	2	4	5	8	8	8	4	9	9	0	4	9	8	4	4	2	1	1	4	22
53	0	2	0	1	4	6	6	4	0	0	0	0	0	3	6	2	2	0	2	4	13
54	6	2	0	1	6	8	4	4	6	2	2	2	6	8	6	4	2	2	0	4	20
55	2	0	0	0	4	3	4	2	0	0	0	0	0	0	0	4	1	2	0	3	17
56	4	0	4	2	6	7	0	2	4	0	9	3	9	8	2	4	0	0	0	0	17
57	–	–	–	–	–	–	–	–	0	0	0	0	0	8	4	4	2	1	0	3	17
58	–	–	–	–	–	–	–	–	2	0	0	4	0	5	0	0	2	1	2	5	15
59	–	–	–	–	–	–	–	–	6	0	0	0	0	8	4	4	1	2	0	3	22
60	2	6	2	6	8	7	4	4	0	4	9	1	9	6	6	4	2	2	2	6	19
61	–	–	–	–	–	–	–	–	0	0	0	2	0	0	4	4	1	2	1	4	11
62	0	0	2	0	0	8	0	4	0	0	4	2	0	5	2	4	2	1	2	5	25
63	2	2	4	3	6	3	4	4	2	0	0	5	0	6	4	4	–	–	–	–	--
64	0	0	2	1	4	5	4	4	0	0	0	0	0	4	4	4	2	2	0	4	7
65	8	0	2	1	6	8	6	4	8	0	0	1	4	4	4	4	2	2	2	6	27
66	0	0	2	1	4	8	4	4	0	0	2	0	0	7	4	4	1	2	0	3	--
67	6	0	2	1	6	6	4	0	2	2	0	0	0	2	0	2	0	2	0	2	9
68	4	2	4	5	4	8	4	4	6	0	0	7	6	6	8	4	2	1	2	5	21
69	6	0	2	4	6	8	8	4	6	6	2	3	9	5	0	2	1	2	2	5	25
70	2	4	2	6	6	8	6	4	6	4	2	3	2	6	2	4	1	1	1	3	15
71	4	0	0	0	6	8	8	4	4	6	0	4	0	4	0	4	1	2	2	5	23

....continued

	Final Test								Quizzes								DL				
[illegible]	A	B	C	D	E	F	G	H	A	B	C	D	E	F	G	H	TS	RT	RA	DL	SK
2	8	2	0	1	8	8	6	4	6	4	0	0	0	6	8	4	2	2	2	6	22
3	6	6	4	6	6	8	8	4	6	6	2	2	8	7	8	4	2	2	0	4	18
4	6	8	4	8	8	8	8	4	9	9	9	5	9	7	8	4	2	2	2	6	20
5	0	0	0	0	0	0	2	0	9	0	2	0	9	5	4	2	1	1	0	2	9
6	0	0	0	0	0	6	4	4	0	0	0	2	0	4	0	4	2	1	2	5	13
7	6	6	8	6	8	8	8	4	4	2	9	2	9	8	6	4	2	2	2	6	21
1	4	2	8	1	4	8	2	4	9	9	0	3	6	4	2	2	–	–	–	–	--
2	8	8	6	8	8	8	8	4	9	9	8	6	6	8	8	4	0	1	2	3	--
3	2	0	2	3	2	6	0	4	9	9	0	0	0	2	4	4	2	2	0	4	17
4	8	8	8	8	8	8	8	4	9	9	8	8	6	8	8	4	2	2	2	6	36
5	6	6	4	7	8	8	8	4	2	4	2	2	2	8	0	4	2	2	1	5	30
6	6	6	8	6	6	8	8	4	9	9	8	4	8	8	8	4	2	2	2	6	29
7	8	8	8	8	8	8	8	4	9	9	0	8	8	8	8	4	2	2	2	6	20
8	8	8	8	8	8	8	8	4	9	9	8	6	6	8	8	4	2	2	2	6	32
9	0	6	8	8	6	8	8	4	9	9	9	6	8	7	4	4	2	1	2	5	39
0	0	2	0	2	6	4	0	4	2	2	0	7	0	6	2	4	2	2	4	6	40
1	8	8	6	8	8	7	8	4	9	9	9	9	9	9	8	4	2	2	2	6	17
2	8	6	8	8	8	8	8	4	9	9	8	8	8	8	8	4	2	2	2	6	41
3	8	8	0	6	8	8	4	4	6	2	9	9	9	9	0	4	2	2	2	6	31
4	8	6	8	8	8	8	8	2	9	9	8	5	8	8	4	4	1	2	1	4	14
5	6	8	8	8	8	8	8	2	9	9	0	2	9	8	8	4	–	–	–	–	--
6	0	4	2	1	4	8	8	4	9	9	9	9	9	9	2	4	2	1	2	5	25
7	4	8	8	8	8	8	8	4	9	9	0	7	4	8	8	4	2	2	2	6	20
8	2	0	4	2	2	5	2	4	9	9	0	0	2	4	4	4	0	1	1	2	14
9	0	4	0	7	2	3	8	4	9	9	9	9	9	9	4	4	1	2	1	4	29
0	0	4	0	8	4	8	4	4	9	9	2	1	9	7	4	4	2	2	0	4	32
1	4	0	0	2	4	4	4	4	6	0	0	2	2	3	2	4	2	1	2	5	20
2	2	4	4	5	2	7	8	4	9	9	9	4	9	8	4	4	2	1	2	5	17
3	6	2	6	5	8	8	4	4	9	9	9	9	9	9	8	4	2	2	1	5	17
4	2	0	2	1	4	8	3	4	9	9	0	0	2	7	6	4	1	1	0	2	14
5	–	–	–	–	–	–	–	–	6	2	2	4	2	8	6	4	–	–	–	–	--
6	4	4	4	6	4	8	8	4	9	9	9	5	9	8	4	4	–	–	–	–	--
7	6	4	8	4	6	8	6	4	9	9	6	4	4	8	4	4	2	2	2	6	14
8	6	6	6	6	8	8	8	4	9	9	9	6	8	8	9	9	0	1	2	3	28

....continued

ID	Final Test A	B	C	D	E	F	G	H	Quizzes A	B	C	D	E	F	G	H	DL TS	RT	RA	DL	SK
529	4	6	0	3	4	8	6	4	9	9	9	5	9	7	9	9	1	2	2	5	18
530	8	8	8	8	8	8	8	4	8	8	8	8	8	8	8	4	2	1	2	5	33
531	4	8	8	8	8	8	8	4	9	9	8	7	8	8	8	4	–	–	–	–	--
532	2	2	6	4	8	7	6	4	4	2	9	9	9	9	0	4	1	2	1	4	20
533	8	4	4	3	8	8	8	4	2	8	2	2	2	8	8	4	1	2	1	4	26
534	8	8	8	8	8	8	8	4	9	9	9	8	9	8	8	4	2	2	2	6	15
601	6	8	8	8	8	7	8	4	8	6	8	6	8	6	8	4	2	1	2	5	21
602	2	4	0	2	8	8	6	4	9	9	8	8	8	8	9	9	2	2	2	6	34
603	6	8	8	8	8	8	8	4	8	8	9	8	8	8	0	4	2	2	1		39
604	4	4	4	8	6	7	8	4	4	2	2	5	6	8	4	4	1	2	0	3	30
605	6	6	4	2	6	8	8	4	8	6	8	5	6	8	8	4	2	1	2	5	33
606	8	8	8	8	8	8	8	4	6	8	9	8	6	7	8	4	2	2	2	6	32
607	6	8	6	8	8	8	8	4	8	8	9	6	8	8	0	4	2	2	2	6	31
608	4	4	0	3	6	5	4	4	2	2	0	2	0	7	4	0	2	1	0	3	11
609	6	8	6	6	8	8	6	4	9	9	4	8	6	8	8	4	2	2	1	5	27
610	3	8	6	8	6	6	6	4	0	4	9	8	9	8	0	4	1	1	2	4	27
611	8	8	8	8	8	8	8	4	9	9	8	8	8	8	8	4	2	2	2	6	27
612	8	4	8	8	8	8	8	4	6	8	8	8	8	8	8	4	2	2	2	6	28
613	8	8	4	8	8	8	4	0	9	9	9	9	9	9	9	9	–	–	–	–	26
614	6	8	2	7	8	6	6	4	6	8	2	5	8	6	6	4	–	–	–	–	18
615	6	6	8	8	9	7	8	4	8	8	8	8	6	8	8	4	2	2	2	6	41
616	8	8	8	8	8	8	8	4	8	8	8	8	8	8	4	4	2	2	2	6	30
617	8	8	0	8	6	8	6	2	6	8	9	9	9	9	9	9	2	2	2	6	19
618	8	8	8	7	6	8	8	4	8	6	8	8	6	8	8	4	–	–	–	–	--
619	8	8	8	8	6	8	8	4	8	8	9	9	9	9	6	4	2	2	2	6	11
620	4	6	8	8	4	7	8	4	8	8	8	8	8	8	8	4	2	2	2	6	24
621	6	8	4	6	8	8	8	4	8	4	9	8	4	8	8	4	2	2	1	5	18
622	0	6	0	4	4	8	2	0	2	0	0	6	0	8	2	4	2	1	1	4	23
623	8	8	8	8	8	8	8	4	8	8	8	8	8	8	9	9	2	2	1	5	28