AN INFORMATION PROCESSING APPROACH TO THE INVESTIGATION OF MATHEMATICAL PROBLEM SOLVING AT SECONDARY AND UNIVERSITY LEVELS

VOLUME 1

BY

MOHAMMED TAHAR TALBI

A thesis submitted in part fulfilment of the requirement for the degree of Doctor of Philosophy of the University of Glasgow, Faculty of Science

ACKNOWLEDGEMENTS

I am most indebted to Professor A.H. Johnstone and Dr. D.B. Webber not just as research supervisors but also for their advice, guidance, patience and inexhaustible enthusiasm during the period of the study.

I am grateful to members of staff and students of the Mathematics Department of Glasgow University and members of staff and pupils of the Scottish and Algerian schools which participated in this work and provided the necessary material to complete it. Many thanks are also due to the SCE Examination Board for offering the data which were essential for my work.

I would like to express my appreciation for all the support I received from my colleagues in the Centre for Science Education at the University of Glasgow throughout my research career.

I would like to thank my Ministry of National Education not just for providing the finance for this work but also for its encouragement throughout it.

Finally, I wish to express my great debt of gratitude to my wife and my children for their encouragement and patience upon which my thesis depends.

CONTENTS

<u>Chapter 7</u>

ABSTRACT

This thesis contains ten chapters: three of them are background literature and five have resulted from practical work during the whole period of the research. Chapter 9 is an attempt to extend the idea of the demand of a task, while the last chapter contains conclusions and suggestions for further research.

In Chapter 1, the theories of Piaget, Gagné and Ausubel are described and compared with each other. Piaget's stages of intellectual development and how learning processes take place are described and explained. The contribution of the theory in the domains of curriculum, teaching Piagetian tasks as subject matter and matching instruction to development stages is stressed. However, the serious challenges to the theory are (i) the horizontal decalage phenomenon, (ii) relating stages with age, (iii) assessing competence and readiness.

Gagné's model of an hierarchy of learning comes from theories of transfer. It is built from the top down. The conditions of learning are internal and external and ranged from signal learning to problem solving. The learning process is based on associational chains. The difficulty of the model comes from the nature of a learning hierarchy and its validation.

Ausubel's theory of meaningful learning is based on what the learner already knows. It is built up from seven elements which range from meaningful learning to the advance organizer. Meaningful learning occurs as a result of interaction between new and existing knowledge and its variation is due to the growth of differentiation and integration of relevant items in cognitive structure. Failure in

learning may occur in situations such as those of conflicting ideas and forgetting.

In Chapter 2, **Information Processing Theories of Learning** are described and the justification of these theories as a fourth paradigm to guide thinking about research is stressed. A model of human memory is given and the components of memory and their features are listed. Stress is placed upon the memory processes and their levels, organization of knowledge, working memory and chunking as a remedy for overload.

Two examples of these theories are given namely **Neo-Piagetian Theory** and the **Predictive Model of Holding-Thinking Space**. The main goal of the former is to make Piaget's theory functional not just structural. The latter relates performance to the amount of information to be processed in learning and problem solving. This model is applied in both University and Algerian samples. This can be found in Chapter 3.

In Chapter 4, the field dependent-independent cognitive style is considered as an important factor affecting performance. The differences between field dependent-independent people may be related to the perceptual field, selected information and the level of guidance. The reason for these differences may be due to the way in which information is both analysed and represented in memory. The practical work has been done with both University and Algerian samples.

In Chapter 5, some other factors are described. Most of them are concerned directly with the subject matter. The activities involved in learning mathematics are classified and attention is given to Polya's version of **heuristic strategies**. The concept of understanding is considered as a basic goal of education and its meaning is given in three different aspects. Most attention is given to the third one, which is known as alternative framework or misconception. The levels of understanding of Skemp are defined and their goals are stressed. The causes of learning difficulties in mathematics are listed, while the different forms of mathematical language are described and their affect on learning is noted.

In Chapter 6, the analysis of Paper I (multiple-choice questions) has been done for preliminary Examination of four Scottish schools (a fifth school used only traditional questions). The experimental work is concerned with language, formulation and type of question. The analysis of Paper II (traditional questions) has been done for preliminary Examination of the above five schools and the SCE Examination. This can be found in Chapter 7.

In Chapter 8, experimental work (concerning Paper II) is described in terms of its material, techniques used, experimental design and how the test was administered. In this experiment, instructions, sample questions and sample grids were provided.

In Chapter 9, a "new" method for assessing the demand of a question is described and applied. The method was devised as a result of difficulties raised in applying the relative demand to the data.

In Chapter 10, conclusions and recommendations are presented and suggestions for further research are listed.

INTRODUCTION

Psychology as it applies to mathematics education is important for several reasons such as:

(i) it offers some insights into the learning and teaching process;

(ii) it develops some broad theories which can be translated into instruction in classroom situations;

(iii) it encourages the teacher to observe his/her subjects systematically and to monitor his/her instruction with more care.

Teaching and learning should be taken as a continuous process. This process can be facilited by looking **"backwards"** and **"forwards"** to find out what the learner already knows, what must be taught and learned, the nature of the subject matter and the educational system.

One of the most important factors which affect both learning and teaching is the limitation in the size of **"working memory space"**. As a consequence of this, the amount of information which can be held in conscious attention at one time is limited too. Therefore, any increasing of this amount may lead to **"overload"** and **"poor learning"**. This can be overcome by developing a **"strategy"** which permits the learner to group and **"chunk"** information in a meaningful way.

Another factor which affects the success and failure in learning and teaching is the **"demand"** of a task in terms of what has to be recalled, transformed, deduced and concluded. The size of working memory space cannot be changed. But the task's complexity can be modified. There is no agreement about how to count its **"thought steps"**. The method of looking at the subjects' scripts may lead to "relative" demand rather than "**absolute**" demand which may involve "factors" not shown in the subjects' scripts.

The practical work of the research is concerned with the identification of some of these factors and the application of them in order to modify the demand of a task. This has been done successfully. The "new" estimation takes into account the "construction" of questions in terms of how their parts link together, the factors which are found to have a significant affect on subjects' performance and the relative demand as it is established in the literature.

It is hoped that from this work, teachers and pupils will benefit from improved learning.

CHAPTER ONE

A REVIEW OF THE RELEVANT MODELS OF LEARNING

PIAGET'S THEORY OF INTELLECTUAL DEVELOPMENT

1. INTRODUCTION

The meaning of "cognition" and "cognitive structure" is the acquisition of knowledge and the organisation of this knowledge in memory. Investigations into whether cognitive structure changes with time or is inborn (or at least developed at an early age) led [1] to developmental and non-developmental theories since the former supports the first assumption, whereas the latter the second. An example of the first category is Piaget's theory, while information processing theories belong to the second category.

2. PIAGET'S THEORY

Piaget has had great influence through his description of childrens' behaviour. The affect of the biological factor in his theory is obvious. He [2] conceptualized intellectual development as acts of adaptation to the physical environment and the organization of this environment. He believed that [3] the fundamental characteristics of human thinking could be understood in terms of the logical propositions and relationships that human behaviour expressed. The direct observation of the thought process is difficult; but activities based on interviewing led Piaget to sequence the geometrical properties: topological, projective and Euclidean. However, this psychological order reverses the historical order [4].

3. LEARNING PROCESS

Piaget made a distinction between physical and mathematical knowledge. This resulted [1] from the difference between the construction of logical structure or "schemes" from empirical generalization or inference made from physical experience. The learning process is a matter of active thinking and of operating on the environment since the interaction with the environment leads to the gaining of new experiences, more learning and more adaptation to it [5]. However, this adaptation occurs [6] through interplay of the processes of assimilation and accommodation: new material is assimilated to existing ones, but if there is too much then a cognitive conflict occurs which is resolved by an accommodation [1].

4. THE STAGES OF INTELLECTUAL DEVELOPMENT

The observations of Piaget of children and adolescents led him to note qualitative differences in the structure and nature of intellectual behaviour [7], the presence or absence of certain operations [3], the modes of reasoning [8] and thinking [3]. These qualitative changes which occur as a result of an equilibration process [5], led Piaget to sequence the stages of intellectual development as: sensory-motor, pre-operational, concrete operational and formal operational. Each stage represents a set of levels of equilibration and all children develop mentally and pass through these stages in the same order but not at the same rate [5]. Only the last stage is relevant to my work.

Formal operational stage

The thinking and reasoning in this stage have been characterized as follows. The thinking is propositional; it involves combinatorial analysis and, at the end of the period's stage, it reaches its maximum. The reasoning is hypothetico-deductive; children at this stage can form and test hypotheses, handle abstraction, use pure symbols and separate variables. In addition to this, assumptions are more readily made, general laws are obtained and quoted, and common principles are understood. In short, the stage is characterized by the reaching of a high degree of equilibrium.

5. THE CONTRIBUTION OF THE THEORY TO THE FIELD OF MATHEMATICS EDUCATION

(a) Curriculum application

In the early 1960s and the early 1970s a considerable number of curriculum projects appeared in mathematics. Each was influenced to some extent by Piaget's work. For example, the Schools Council/Nuffield Project was designed in the light of a Piagetian view, and its materials [9] were heavily influenced by Piaget's work. The Concept in Secondary Mathematics and Science project [10] tried to identify different levels of cognitive functioning and hierarchies of understanding for a number of topics. The Graded Assessment in Mathematics Project [11] also adopted the idea of Piaget's learning hierarchy. The Cockroft Report has no great enthusiasm for Piaget's stages or for any views on readiness [7].

In short, this period may be characterized by the growth of mathematical projects, the analysis of curricula and materials into their concrete and formal thinking in order to determine the demand of an item and then to provide a hierarchy in the teaching and learning process.

(b) Teaching Piagetian tasks as subject matter

In order to accelerate teaching and learning processes, many experiments have been carried out [8]. Some of them had a little success on retention and transfer, others induced improvements in childrens' cognitive performance. However, the results of Siegler (1973, 1975) [12, 13], Case (1978) [14] and Donaldson (1978) [15] with young children show, in a clear manner, the weakness of Piaget's opinion that one must wait for childrens' readiness for formal operations training.

(c) Matching instruction to development stages

The idea of a "matching model" requires that [3] both content and presentational techniques of teaching should be matched to the child's current level of development. To achieve that, one should [16] identify the appropriate stage reached by an individual and analyse the curriculum tasks for their level of cognitive demand. Meaningful learning will occur only when the cognitive skills demanded by the task are available to the learner. However, this leads to the problem of "readiness" in which teachers should wait until their pupils are ready.

The general principles which derived from Piaget's theory and which may help educational procedures are constructive learning, concrete representation, social feedback and clinical teacher-pupil interaction.

6. CRITICISMS

(a) Piaget's stages

There is considerable evidence that individual children cannot easily be categorised as being at a particular stage of development. Therefore, the "horizontal decalage" (the phenomenon of passing certain tasks and failing others with the same logical structure) represents a serious challenge to a strict stage theory [17].

The "experience with mathematical tasks" as opposed to "more generalized experience with the environment" is more likely to improve children's ability to apply logical structure [3].

The explanation of mathematical understanding in the light of Piagetian stages theory was found difficult in the work of the Schools Council Project [18], the Concept in Secondary Mathematics and Science Project [19] and the Strategies and Errors in Secondary Mathematics Research [20].

The order of acquisition of mathematical concepts contained in the Nuffield "Checking Up" books is generally not confirmed through cross studies (e.g. Brown [21] & Orton [7]).

The problems caused by attempting to relate stages with age are very well known. It is difficult to identify categorically at which stage a particular pupil is at a particular moment in time. El-Banna [5] reported that the age difference may be as much as ±2 years, while in some difficult topics (such as place-value skill) it may be even bigger. Orton noted that a proportion of the population never develop those abilities outlined by Piaget as being characteristic of formal operational thinking. The nature of formal operation is not clearly defined and it is difficult to apply it to examples of mathematical operations [21]. This lack of an agreed definition of formal operational thought led Jenkins [22] to conclude that it is very difficult to define the level required to understand a particular topic in a school course.

<u>(b) The difficulty of assessing competence</u>

From Piaget's point of view, failure in a task is caused by a lack of competence, but the empirical evidence of many researchers (e.g. Resnick [3] & Donaldson [15]) shows that the failure could depend on a number of separate variables like knowledge, language, display hierarchy etc. Donaldson doubted that failure in Piaget's tasks show evidence of failure to decentre and failure to reason. The lack of familiarity, and the complexity of instructions have been mentioned by Orton [7] who stated that the unusual and unexpected, even unacceptable, nature of the question (posed by Piaget) might have seriously influenced the results.

<u>(c) Readiness</u>

The weakness of the position "children must be biologically ready" is very clear since there is much evidence in the field which contradicts it. Bruner [23] stated that we begin with the hypothesis that any subject can be taught effectively in some intellectually honest form to any child at any stage of development. Andersson [24] regarded a person's thinking as "local" and content related rather than as a result of the fact that a given individual finds himself at a certain stage of thinking.

To overcome the readiness difficulty Hunt [25, 26] proposed that the important thing in education is always to pose problems that are slightly beyond the learner's current capability but not so far beyond that they are incomprehensible. This position has been affirmed by Donaldson [15] who stated that you cannot master any formal system unless you have learned to take at least some steps beyond the bounds of human sense.

(d) Sampling

Piaget's experiments were tried on a small number of children, therefore his work has not always satisfied the requirement of scientific research [2]. The work with a small sample size must be treated with caution, but the information obtained is not necessary invalid [7].

Conclusion

Maybe the opinion of Pascual-Leone [27] gives a very clear picture of the Piaget theory: "Piaget's theory is a competence model since it defines the ideal behaviours for each stage, and it does not provide an application of how the content of mental operations are selected, organized or sequenced or how performance characteristics, such as memory or attention, limit the child's responses." Much of the intellectual framework which has been applied in schooling is just misleading since, as Donaldson confirmed, rational thinking by children certainly will occur if we provide appropriate material in a right way and a right language.

GAGNÉ'S MODEL OF LEARNING HIERARCHY

1. INTRODUCTION

In our introduction to Piaget's theory, a difference between developmental and non-developmental theories has been noted. However, Sharrat [1] gives other kinds of developmental theories which are based on "assumptions of continuity, with changes in performance attributed to units added through experience." He noted that the continuity theories are best represented by Gagné's neobehaviourist theory of cognitive learning.

It was noted [28] that the successful training programs for military personnel during and after World War II and the careful sequencing of such programs helped the learner master prerequisite skills. This success promoted hierarchical learning.

According to Resnick [3], the "identical elements" theory of transfer supports the notion that successfully learning one task would make it easier to learn a second task to the extent that the two tasks contained some of the same components (i.e. the same sets of associations).

The study of a variety of theories of transfer led Gagné to observe that certain experimental situations serve as typical models of learning such as trial and error learning, conditioned responses, verbal associations and studies of insight. Hence, he [29, 30] tried to convert these theories of transfer from their laboratory study to something more realistic such as the school curriculum. As a result of this a "cumulative learning theory" emerged which is a special version of an identical elements theory [3].

The Gagné model [31] was not intended to describe a theory of learning but rather to make a bridge between the laboratory and classroom learning.

2. LEARNING THROUGH THE MODEL

(a) How the model was built

It was noted [7] that Gagné's model of a learning hierarchy was built from the top down. At the top we define the final capability required to accomplish a task in terms of behavioural objectives (i.e. the ability to do such things). The next stage is analysing the task by considering what prerequisite capabilities are required in order to be able to show the final capability. However, moving from one stage to the next, one needs to ask the question "what would one have to know or do in order to perform this task, after being given only instructions?" [32]. The answer would be the prerequisites of the target task and each of them may have its own prerequisites in turn. These processes can be repeated until a complete hierarchy of successively simpler skills is generated [3].

(b) The conditions of learning

Because there are different types of learner capabilities, one may expect several varieties of performance types which may also be differentiated in terms of the conditions of their learning [5]. These conditions were classified by Gagné as internal and external. The former, previously learned capabilities, were found to be better predictors of student achievement than other indicators of student abilities such as grades in previous courses and general intelligence.

The latter were predictable and controlled manipulations of the learning environment around them [28].

In his book "The Conditions of Learning", Gagné [30] proposed eight types of learning. These types are ranged according to their complexity from signal learning to problem solving:

(1) signal learning,

(2) stimulus-response learning,

(3) chaining,

(4) verbal association,

(5) multiple-discrimination learning,

(6) concept learning,

(7) principle (or rule) learning,

(8) problem solving.

It was noted [1] that Gagné's descriptions of the first five types are all based on stimulus-response theory, while rule learning (involving combinations of concepts) is described as an "inferred capability that enables the individual to respond to a class of stimulus situations with a class of performance". Problem solving is "not simply a matter of application of previously learned rules" but also a "process that yields new learning". It is clear that problem solving is the highest ability and all other types are prerequisites of it.

A model of mathematical learning should [21] distinguish between four aspects which are:

(1) simple recall,

(2) algorithmic learning,

(3) conceptual learning,

(4) problem-solving strategies.

It was noted that society at large tends to identify achievement

in mathematics with attainment in the first two aspects. Teachers mainly concentrate on the second and third, and educationists value especially the third and fourth. However, these aspects are presented in Gagné's classification under the headings:
Stimulus-response learning (the first five types of the list), rule learning, concept learning and rule learning, and problem solving respectively.

The process of learning, in Gagné's view, has been described by many researchers (e.g. Orton [7], Sharrat [1], Resnick [3], etc.). The learning process is based on associational chains: knowledge gained from new experience becomes associatively linked with old knowledge. This link, therefore allows knowledge to accumulate and work together in the learning of new skills. However, according to this, children learn an ordered, additive set of units of knowledge (or experience). Each new unit is more advanced that the prerequisite units on which it is built. This view suggests that:

(a) a learning hierarchy starts from the skills which the learner already has (his/her prior-knowledge);

(b) a child is ready to learn a unit if he has mastered its prerequisites; in other words, a learner's ability to master high levels of learning is dependent on his bringing prerequisite knowledge and skills to the learning task;

(c) as a result of the above, a failure to perform the complex task can be traced to a lack of competence in one or more of the subtasks.

3. DIFFICULTIES OF THE MODEL

Applying Gagné's model to the teaching and learning process may raise some difficulties which relate both to the nature of hierarchical learning and to its validity.

(a) Criticisms of the Gagné learning hierarchy

In Gagné's learning hierarchy, the more advanced kinds of learning can take place only when a person has mastered a large variety of verbal associations [33]. Therefore, the recall involved in using the hierarchy would soon prove to be a gross memory overload [34].

Because Gagné defined "intellectual skills" behaviourally and he distinguished them from factual knowledge (memorized number facts or general understanding of mathematical structures and relations), then their procedural components stand in certain relationships to each other but the organization of general knowledge underlying these procedures can be very different [3].

The nature of a learning hierarchy suggests that the subordinate tasks are components of the highest level tasks. Therefore, the abilities required early in learning may influence later learning. However, this is not the only way since several kinds of transition relationships that may be needed to account fully for the development of competence in cognitive tasks have been suggested [35].

The higher level tasks in the hierarchy are indeed more complex than the ones below, but this does not mean that they are harder to learn or that it will require more time and effort than each of the lower level tasks did. It was reported that the highest-order skill in the hierarchy may be easy to learn once all its components have been learned [36]. However, Dienes [37] reported that a student who learned a mathematically more complex game first, learned a simpler version of the game more quickly than those who learned the simpler game first. He [38] suggested that certain mathematical concepts may need to be introduced in some measure of complexity rather than in the small sequential steps suggested by analysis into simpler components. Nevertheless, other researchers (e.g. Caruso & Resnick [39]) noted that most students learned best when the skills were taught in the hypothesised hierarchical order, but that a few were able to learn higher-level components without first learning the lower-level ones.

(b) Validity of the Gagné model of learning hierarchy

Much work has been done by Gagné and his colleagues and other researchers (e.g. Wang et al. [40], Gelman & Gallistel [41], Gagné et al. [42], etc.). This work has been concerned with whether the hypothesised prerequisites were necessary and sufficient. It was noted [28] that in a valid hierarchy, by necessity, most learners who are unable to demonstrate prerequisite skills, should not be able to demonstrate a superordinate skill.

In their comments about a training validation study carried out by Gagné [32], Resnick and Ford [3] stated that although this study lent support to the hierarchy under test, it was far from a tight validation of the detailed sequences of learning that the hierarchy hypothesised. They noted that if transfer has been measured at a more detailed level, then more confidence about this validation is possible.

Another experimental study was carried out to investigate the effects of some variables on the acquisition of knowledge of "elementary non-metric geometry". Gagné [42] stated that the implication of such a finding is that one can affect the efficiency of the learning process quite readily by manipulating the content and sequence of material, but not at all readily by manipulating the repetitiveness and temporal spacing of this content.

According to Orton [7], to carry out a validation at all levels of a hierarchy is very time consuming and things do not always work out perfectly in education. Another problem stressed by Orton is that there seems little likelihood that tightly defined and tested learning hierarchies can be defined for all topics which might at some time be taught in mathematics.

The difficulty of the validation of a hierarchy is stressed by Jones and Russell [28] since hypothesised learning dependencies must be tested between each stated intellectual skill, and in hierarchies consisting of large numbers of skills, this demands numerous comparisons. These difficulties have also been stressed by many workers. For example, White [43] has criticized the model of validation conducted by Gagné on several grounds. His criticisms briefly are:

(1) the hierarchy needs a common sense validity which can be obtained by having it checked by experienced teachers of its subject matter before it is investigated empirically;

(2) in a small sample, it may be quite likely that none of these people are chosen, merely by chance;

(3) imprecise definitions lead to imprecise tests;

(4) using only one question for each element prevents any estimate of the reliability of the assessment of subjects;

(5) the delay in testing possession of the elements can lead to rejection of valid hierarchical connections;

(6) the model's validation does not cover previously overlooked connections;

(7) the model lacks an objective way of determining whether or not the numbers are too large for the connection to be accepted as valid.

White proposed a new model of validation in order to overcome these above criticisms.

4. IMPLICATIONS FOR TEACHING

Many workers produced materials and used them in teaching and learning hierarchies (e.g. Gagné et al. [42], Resnick et al. [44], Trembath & White [45], etc.).

Gagné and his colleagues employed "learning programs" in a variety of mathematical topics. They reported that the most prominent implication of this study is that acquisition of new knowledge depends upon the recall of the old knowledge; therefore, the design of an instructional situation is basically a matter of designing a sequence of topics.

Resnick et al. designed carefully an introductory mathematics curriculum and used it as an individualised instructional program. They noted that this method of teaching helps to ensure that every child is given the best chance of learning successfully and that a hierarchy provides a structured sequence for the teacher and student but not a completely determined one.

Trembath and White obtained a hierarchy by analysing students' errors on a previous test and used it for teaching. They reported that the one hour's instruction produced results superior to those of pupils three years older who learnt the topic, as part of their normal curriculum, over a considerably greater time.

Taking into account these and other experiments and the difficulties of the learning hierarchy (such as overload and validation), one can assume that we should always interpret the details of learning hierarchies with caution.

The implication for teaching and learning may be that: it is helpful to use a hierarchy as a map for a sequence of instruction. This may be attained through lesson planning and presentation since the careful sequencing of materials to be learned is likely to enhance the quality of learning.

The "breaking down" into components is not simple enough to cope with and, in addition, the confusion will be cumulative if the teacher does not check whether the objectives for each level in the hierarchy have been mastered before moving to the next one.

To sum up, learning hierarchies can be useful tools to help teachers and instructional designers make explicit their understanding of the organization of skill learning and the way individual children differ in the extent of their learning. But caution is necessary in developing these hierarchies and flexibility is required in using them for ensuring that all children master the essentials of school mathematics, especially computational skills.

5. COMPARISON BETWEEN PIAGET'S THEORY AND GAGNÉ'S MODEL

The comparison between Piaget's theory of mental development and Gagné's model of learning hierarchy has been made by many workers (e.g. Sharrat [1] & MacGuire [46]). The main points are related to the learner, teaching style and material.

Piaget's theory is based on the assumption that "internal organization" is a scheme (or structure), but the Gagné model is based on an associational chain (or link). However, the state of the learner in the former is a function of the stage of structural organization, whereas in the latter is a function of the learned hierarchy of skills.

Piaget has stressed the development levels of cognitive ability, while Gagné emphasised the importance of prior knowledge in providing further learning. However, both agreed that the development status of the learner is a significant factor in determining his ability to learn but they differ in the nature of this development.

In teaching, both emphasised the importance of correct sequencing but the "Geneva school" favoured the conflict cognitive method, whereas Gagné prefered an expository teaching style.

Finally, Piaget does not distinguish different types of material but does specify different modes of learning in terms of reasoning patterns available at different stages of intellectual development.

AUSUBEL'S THEORY OF MEANINGFUL LEARNING

1. INTRODUCTION

The theory of meaningful learning proposed by Ausubel [47] was a general theory and was not specific to mathematics. Therefore, most mathematics educators have not paid much attention to this theory [7]. It is unusual to find mathematical references dealing directly with this theory. However, I believe that a general theory of learning may have as much to offer as any specific theory since, on one hand, the isolation of mathematical ability from other abilities is difficult and on the other, a theory of learning which is based on real classroom situations, which takes into account the structure of the subject matter and what the learner already knows, and recommends the preparation of the cognitive structure of the learner to accept a new idea, should be very welcome.

2. AUSUBEL'S THEORY OF MEANINGFUL LEARNING

The most important factor in Ausubel's theory is the prior knowledge of the learner. He [47] stated that "If I had to reduce all of educational psychology to just one principle, I would say this: the most important single factor influencing learning is what the learner already knows. Ascertain this and teach him accordingly." To understand what "prior knowledge" of the learner means from Ausubel's point of view, we must understand the components of his theory which are as follows [48]:

(1) meaningful learning versus rote learning,

(2) subsumption,

(3) obliterative subsumption,

(4) progressive differentiation,

(5) superordinate learning,

(6) integrative reconciliation,

(7) advance organiser.

I will deal with these components in some different degrees of detail, trying to illustrate most of them by mathematical examples.

In this theory, there are [46] two fundamental, independent dimensions of the learning process: the information presented to the learner by reception or by discovery, and assimilation of this information into his existing cognitive structure by meaningful or by rote learning.

1. Meaningful and Rote Learning.

Meaningful learning is an active process of transferring new knowledge to the existing knowledge in the individual's cognitive structure. However, Ausubel's description of this process is that [49] meaningful learning takes place if the learning task is related in a non-arbitrary and non-verbatim fashion to the learner's existing structure of knowledge. According to this view [7], if new knowledge was assimilated within an existing knowledge structure as a related unit, and if appropriate modification of prior knowledge (accommodation) had taken place, the result was meaningful learning.

In short, the meaningfulness of learning should involve interaction between the new and existing knowledge. This may happen if the following conditions of learning proposed by Ausubel are achieved:

(a) cognitive structure has relevant items already there;

(b) new material is logically related to what is there;

(c) the learner is disposed to relate ideas in this way [46].

Rote learning occurs if at least one of these conditions is not met. In this situation, the interaction between new and existing knowledge cannot take place since the new knowledge would have to be learned by rote and stored in an arbitrary and disconnected manner [7].

The nature and degree of differentiation of relevant items vary from learner to learner [50]. Therefore, the meaningfulness also varies. This variation led Ausubel to suggest [5] four kinds of meaningful learning:

(i) concept formation,

(ii) concept assimilation,

(iii) proposition,

(iv) discovery learning.

2. Discovery and Expository (or Reception) Learning.

Expository learning is the situation in which the material to be learned is presented completely to the learner, while in discovery learning, the learner should identify some of this material independently [46]. Ausubel was in favour of using expository methods rather than discovery methods and he justified his position by the following:

(a) guided discovery only looked best because of what it had been compared with, usually rote learning;

(b) there was just no evidence that discovery of any kind was a more effective teaching method than meaningful exposition [7];

(c) most concepts are learned by concept assimilation rather than concept formation [49].

But Ausubel [7] agreed that discovery is important in promoting learning with young children and Novak [50] noted that discovery learning occurs primarily with very young children in the process of concept formation. However, Ausubel also accepted that discovery has a place in the learning of generalisations. This opinion is indeed important since a large part of mathematical learning consists of generalisation. An example of a mathematical generalization which may be discovered is that:

"The sum of the angles of a triangle is 180°."

This could be discovered empirically (by measurement) or by deduction (showing that the sum equals a straight angle).

In contrast to Ausubel's opinion is Bruner's who was the main advocate of discovery learning in the USA around and before 1970 [7]. Bruner [51] favoured learning mathematics by discovery because:

(a) discovery encouraged a way of learning mathematics by doing mathematics;

(b) it encouraged the development of a view that mathematics was a process rather than a finished product.

Bruner noted that because one could not wait for ever for pupils to discover and the curriculum could not be completely open and some pupils might even find their inability to discover extremely discouraging, so discovery should be, to some extent, guided.

Teaching by discovery or by expository methods has obtained a great deal of attention since much work has been done concerning this theme (discovery/expository learning) in mathematics. The general findings are [8]: discovery is often less effective than exposition for immediate learning, but it is better for retention and for transfer to new situations. However, the commonly found advantages of discovery learning in the light of research [8] are:

(a) it ensures meaningful learning, since the pre-requisite knowledge must be activated before the discovery activity can progress;

(b) it presents situations in the same ways as those in which the learning will need to be used subsequently;

(c) it promotes the learning not only of the principle itself but of general strategies for the investigation of problems;

(d) if the discovery is successful, it is highly motivating.

3. Readiness.

The existing part of knowledge which the learner already knows was called by Ausubel the subsumer. This part is an anchorage to which the new knowledge is to be linked. If the subsumers or anchoring ideas or concepts were there (in the pupil's cognitive structure) the pupil was effectively ready [7]. Therefore, the Ausubel view of readiness is closer to that of Gagné rather than that of Piaget. It was reported [52] that "Ausubel was in fundamental agreement with Gagné in that the key to readiness was prerequisite knowledge."

During the process of subsumption both the anchoring concept and the new knowledge are modified but retain their separate identities [46]. As a result of the continuity of modification and elaboration in the learner's cognitive structure, meaningful learning occurs. The growing of differentiation and integration of relevant items in the cognitive structure is the reason for the variation in meaningful learning, not the general stages of cognitive development as Piaget claimed [2]. This may explain the fact that the variation in achievement may depend on the specific learning experience rather than on maturation since older children are generally capable of solving more complex (abstract) problems than younger children not because they have some unique cognitive capacity (structure) but rather because the overall level of differentiation and integration of their concepts is much more elaborate [48].

4. Advance Organizer.

The idea that preparing the cognitive structure of the learner for the new learning task will facilitate the learning of this task, came from Ausubel's theory. This may be done by many ways. For example, teacher's questions or curriculum sequence may do the job. This also can be done by using an advance organizer.

In Ausubel's view, even if the child is not ready in the sense of having appropriate subsumers, there is the possibility of using an advance organizer to bridge the gap. He [53] defined it as "more general, more abstract and more inclusive than the ideas and knowledge which were to follow." It is rare to find mathematical examples which satisfy this criterion, but the ingenious method of Matthews (see Orton [7]) for introducing matrix multiplication by using matrices to send and decode messages is one of these examples. Matthews used this method to produce an anchoring concept onto which to link more important applications of matrix multiplication.

In teaching division of fractions, it was found that [54] the concept of inverse operation serves as a stronger advance organizer than the organizers imbedded in either the common-denominator or complex-fraction processes. An example for each one was given:

$$\frac{2}{3} \div \frac{4}{5} = \frac{10}{15} \div \frac{12}{15} = \frac{10}{12} \qquad \text{(common denominator);}$$

$$\frac{2}{3} \div \frac{4}{5} = \frac{\frac{2}{3}}{\frac{4}{5}} \cdot \frac{15}{15} = \frac{10}{12} \qquad \text{(complex fraction);}$$

$$\frac{2}{3} \div \frac{4}{5} = n \quad \text{means} \quad \frac{2}{3} = n \cdot \frac{4}{5},$$

$$\frac{2}{3} \cdot \frac{5}{4} = n \cdot \frac{4}{5} \cdot \frac{5}{4} = n \qquad \text{(inverse operation).}$$

The major attributes of advance organizers have been given by Tamir [50]:

(1) they are not part of the learning material itself;

(2) they are designed to match the prior knowledge of the learner;

(3) they are presented in advance of the learning material;

(4) they present highly inclusive ideas which are capable of creating the anchorage for the subsumption of the specifics of the learning material.

The main goal of the advance organizer is to bridge the gap between what the learner already knows and what he needs to know. But when there are no relevant items in the learner's cognitive structure then, as Novak [48] stressed, it is unlikely that any type of advance organizer will function, for the organizer itself must be meaningful to the learner.

A number of experiments on advance organizers have been done in different subject areas, and the results in general have been mixed (e.g. Ashlock & Waynel [8], Barnes & Clawson [55]).

To sum up, the idea of advance organizer is too useful to be rejected even though the hierarchical nature of mathematics appears to suggest that there should not be many occasions when new knowledge cannot be linked to existing knowledge [7]. But the use of an organizer in the mathematics classroom should be explored. However, the use of a higher level organizer may not be possible without first teaching this more abstract set of concepts.

5. Superordinate and Subordinate Learning.

The organization of knowledge in the mind involves movement or rearrangement of concepts such as gathering and scattering them. Novak and Gowin [56] described this constant movement as "a pushing and pulling of concepts, putting them together and separating them." This may involve the realisation that certain ideas are all part of a more inclusive or superordinate concepts structure [7]. The idea of distinguishing between primary and secondary concepts is discussed by Skemp [57] in the sense that the former is derived from our sensory and motor experiences of the outside world and the latter is abstracted from other concepts.

In Ausubel's view [7], in superordinate learning, the previously learned concepts are seen as elements of a large, more inclusive idea, while in subordinate learning (or progressive differentiation in learning) the most inclusive elements of a concept are introduced first and then the concept is dissected in terms of detail and specificity.

In mathematical classroom situations, the two kinds of learning exist. In fact Orton [7] stated that the kind of reorganisation of knowledge involved in learning mathematics is certainly likely to involve the two way process of relating concepts both to subordinate and to superordinate concepts. He illustrated them by symmetry and binary operation respectively.

The identification of superordinate and subordinate concepts is not only difficult but also complete agreement about them is unlikely. According to Orton [7], Novak noted that determination of what in a body of knowledge are the most general, most inclusive concepts and what are subordinate concepts is not easy. He suggested that a concept map should play a role in curriculum planning which attempts to analyse the relationships between concepts. In addition, concept mapping could facilitate the implementation of attainment targets and also graded assessments [1].

6. Conflict and Failure in Learning.

Many situations in learning have had much attention; for example, when conflict occurs, also when learning either does not take place or is quickly forgotten. The conflict of meaning or cognitive dissonance in the terminology of Ausubel, may occur on many occasions such as using a meaning of an idea which is in conflict with a previously understood idea. This tends to create a disequilibrium. The conflicting ideas create a problem of accommodation which needs a reconciliation.

Ausubel's theory provided a solution for this phenomenon through the process of integrative reconciliation. Without integrative reconciliation, it is possible that learners might compartmentalise the conflicting ideas [7].

There is another factor which affects learning: it is the issue of forgetting. According to Ausubel's theory, the degree of meaningfulness of the learned material can be explained in terms of forgetting rate since, in rote learning, the forgetting may happen sooner except in the case when there is overlearning such as repetition, revision, etc. In the meaningful learning, the retention is certainly much longer but forgetting still occurs because of obliterative subsumption.

An example of obliterative subsumption was given by Orton [7] to show how valuable knowledge may be obliterated without the learner becoming deprived. In one method to factorise the quadratic $10x^2 + 23x + 12$, you need to find two numbers such that their product is 10×12 and their sum is 23. By trial and error, they could be 15 and 8. Then

$$\begin{aligned} 10x^2 + 23x + 12 &= 10x^2 + 15x + 8x + 12 \\ &= 5x(2x + 3) + 4(2x + 3) \\ &= (5x + 4)(2x + 3). \end{aligned}$$

With experience, a pupil can factorize by inspection and may forget the above and other methods without being disadvantaged.

3. CRITICISMS

It is clear that what the learner already knows is of central importance and the educators' agreement about this may be general, but to know in full detail the prior knowledge of the learner is not easy. Therefore, accepting verbal or expository learning as an effective and efficient method for teaching mathematics is quite hard too. Furthermore, it was reported [58] that the obvious relation of Ausubel's theory to the teacher's task make it eminently worthy of consideration and deserves wider acceptance than any other theory. However, this theory is less supported by data since it is experimentally difficult to investigate.

4. CONCLUSION AND COMPARISON

Understanding the learner as an individual and using progressive differentiation and integrative reconcilation in teaching and learning is the core of Ausubel's theory. However, Piaget was concerned with cognitive development and not with individual learning. His interviewing technique is certainly efficient in diagnosing childrens' ideas. This may help to identify the logical structures that enable differentiated concepts to be related to one another and be progressively differentiated [50].

Ausubel accepted [1] the ideas of assimilation and accommodation and referred to concrete and formal (or abstract in his terminology) stages but he does not accept the full implications of Piagetian stage theory.

In terms of learning hierarchies, Ausubel has a similar view to that of Gagné but he takes a more extreme position than Gagné about content knowledge rather than learning to think [59]. However, Ausubel adds to the learning hierarchy the principle of the advance organizer.

In terms of readiness, Ausubel's view was nearer to that of Gagné than to that of Piaget.

In terms of transfer, Ausubel interpreted learning as the continual modification and amendment of the learner's cognitive structure. Piaget has supported the development levels of cognitive ability, while Gagné suggested the association link in providing further learning. In short, the understandability of learned material, the learner's adaptation of meaningful material and the harmony of knowledge in the cognitive structure may play a critical role in anchoring successful teaching and learning.

CHAPTER TWO

INFORMATION PROCESSING THEORIES OF LEARNING

INTRODUCTION

The development of information processing theories began in the 1950s. With the arrival of the electronic computer and its advanced technology to handle more complex reasoning tasks and the greater tendency to accept that the brain may be considered as a processor of information, the assumption that the computer could serve as a model of the brain is quite reasonable.

Barber (1988, [60]) noted that a model of human performance which could be formulated in the light of the learner's abilities may serve as a device for summarising them in a convenient framework. The contribution of the electronic computer to education has been stressed by Orton [7]: "contemporary theories of human learning have frequently looked to the computer as a model of the human mind." However, this idea of analogy has been taken further. These theories, according to Orton, suggest that "the human mind has a built-in ready for action ROM (read only memory) from the moment of birth." Sharrat [1] has related these approaches to "nondevelopmental models" since in general they make the assumption that "cognitive processes or structures do not undergo developmental change." Such processes are either inborn or develop at such an early age. As a result of this analogy between the computer and the brain, a body of learning theories has arrived on the scene, developed by communications engineers and adapted by psychologists to interpret the learning process using computer terminology: "data is received, processed, stored and output" [61]. The process may involve coded information which has to be decoded and transformed into a response. According to Macnab and Cummine [61], information processing attempts to apply the concepts of computer storage, processing and retrieval to the working of the mind. While Sharrat [1] reported that the computer acts as a sort of metaphor to describe general processing mechanisms.

The justification of information processing theory as a fourth paradigm to guide thinking about research in science education was asserted by Stewart and Atkin [62] who noted that although the research of Ausubel, Gagné and Piaget have received a great deal of attention in the literature of science education, each one has been criticised. The information processing view of memory, learning and problem solving encompasses all three of the above paradigms and, to a large extent, overcomes their weaknesses. Some of these weaknesses briefly are [62]:

(1) Gagné's hierarchical learning model has been faulted because of its limited scope (behavioural objectives may be suitable to learning outcomes in the case of skill learning but not other types of learning such as propositional and conceptual learning);

(2) Piagetian researchers often play down the role of prior knowledge in determining performance (a shortcoming of Piaget's theory may be that it lacks specification of detailed mechanisms competent to generate the phenomena it describes);

(3) Ausubel's concept of meaning is unnecessarily vague (a shortcoming of Ausubel's theory may be that it lacks concern

with memory processing mechanisms since the lack of success in learning or problem solving could be found in individuals with well organized cognitive structures who, because of deficient or absent routines for manipulating that information, have trouble solving problems).

However, the more detailed criticisms of these theories appears in the previous chapter during the presentation of each theory.

Resnick [3] looked at information processing theories in the light of structural knowledge in mathematics and thinking: can these theories shed any light on how people understand mathematical concepts, suggest any organization of teaching to promote conceptual understanding and give insights into relationships between conceptual understanding and performance of routine mathematical tasks in order to bridge the gap between conceptual and computational approaches to mathematics instruction? He noted that this bridging is only beginning to be realised but he assumed that "for the first time psychology has a language and a body of experimental methods that is simultaneously addressing both the skills involved in performance and the nature of the comprehension underlying that performance."

Although there are different approaches in information processing theories, the things they have in common are ([1], [3] and [60]):

(1) they attempt to achieve a high degree of precision in describing cognition;

(2) they have power to explain human thinking in describing thought process in terms of symbol manipulation;

(3) they focus on the structure of knowledge within the mind and on the mechanisms by which knowledge is manipulated, transformed and generated in the process of solving

problems;

(4) they explore the manner in which **algorithms** (i.e. systematic solution procedures) and **heuristics** (i.e. procedures for limiting search) enter into problem solving;

(5) they offer the opportunity of assessing human performance through the description of general patterns of success and failure at different stages in the development of a given task.

A MODEL OF HUMAN MEMORY

The traditional laboratory experiments on memory are concerned with the mechanisms of memory, whereas the contemporary view is focused on the memory content in the sense of what is remembered and what is forgotten [63]. In fact Howe [64] used the term memory to denote the capacity of remembering. However, Child [65] regarded memory as "a place, located in the head, where recoverable experience and knowledge are housed." This view is shared by Roth and Frisby [66] who considered the memory as a repository where everything is stored that we need to know to interact with the environment.

Many researchers (e.g. Stewart and Atkin [62], Roth and Frisby [66]) have often used the conceptualization of memory to describe the activities of registering, storing, transferring and retrieving knowledge for further use in learning and problem solving. According to Child [65], the processes hypothesised in contemporary views about memory are encoding (the process of putting information into memory), storage (the methods of retention of information in memory) and retrieval (the process of recovery of stored information from memory). Let us now consider the components of memory, memory process and the organization of knowledge in memory.

THE COMPONENTS OF MEMORY AND THEIR FEATURES

In contrast to the traditional stimulus-response theory, the information processing theory supported the idea that "there were functionally distinct processing mechanisms associated with different classes of memory phenomena" [67]. This opinion came from evidence in the field of memory since many researchers' findings supported the existence of different types of stores in the brain (e.g. [68], [69]). As a result of this, the memory store can be viewed as a series of stages including sensory information storage, short-term memory and long-term memory. In addition to this, there are [62] several processes that ensure the flow of information between the three stores.

It is very important to regard [70] memory as an integral part of the whole processing system and then an item can be processed at different levels within the system. This view helps to overcome the problems of the duplex of memory. According to Craik and Lockhart, the view of memory as being composed of distinct stores with clearly differentiated properties is no longer adequate. For them, the alternative is to focus on how different ways of attending to material affect the degree to which the material is memorised.

SENSORY MEMORY

The main interest in perception and attention models is how we perceive and attend to input, whereas in sensory memory the interest is in how this input kept in order to transfer it into short-term memory [66]. Therefore, the sensory memory refers to the very short period of time in which information is held and registered by our senses. The forms of this information are visual (or iconic) and auditory (or echoic). Maybe these forms reflect the very transitory nature of sensory memory. According to many workers (e.g. Stewart and Alkin [62], Ashlock et al. [8] and Child [65]), the main features of sensory memory are as follows:

(i) material fades in a matter of 0.1 to 0.5 seconds and then is quickly lost unless it is retained within this time and is transfered to short-term memory;

(ii) material is much more accurate, complete and detailed than the material which is stored in short-term memory;

(iii) because there are no rehearsal capabilities (repetition of material in order to be recalled), there is no "instant replay" feature from sensory memory store;

(iv) the corresponding process of attention and selective perception ensures that only particular stimuli are conveyed to the short-term memory.

SHORT-TERM MEMORY

Researchers in the field of memory indicated that the material held in short-term memory disappears in a matter of seconds unless it is kept in consciousness. The maintenance of this material depends on several factors such as the degree of attention we give to it, the usefulness and update of it and on developing some strategies such as rehearsal or repetition which help to transfer and establish this material in long-term memory.

It was noted [65] that rehearsal is closely related to rote learning, the more this cycle is repeated, the more likely it is that information will pass into long-term store. However, this technique depends on the nature of the material and its encoding form.

The limitation of memory span (the capacity of short-term memory) has been established by Miller [71] and it is estimated to hold between 5 and 9 chunks (or items). This has strong support in the research field. Nevertheless, this limitation may explain the small amount of information we can hold and the ease with which it is forgotten by rapid decay or by being displaced by incoming items, since the interference factor also seems to affect short-term memory [66]. However, Reed (1988, [72]) reported that evidence suggests that interference, rather than decay, is the primary cause of forgetting.

To sum up, some features of short-term memory are as follows:

(i) it is a transient information store since it keeps materials for a matter of a few seconds to a few minutes;

(ii) it includes the immediate interpretation of events;

(iii) it is considered as a site of information storage while one tries to organize and store information in long-term memory;

(v) it allows information to reach this level more or less automatically.

LONG-TERM MEMORY

Long-term memory has been divided [73] into semantic and episodic memories. The former refers to the person's general knowledge and the latter refers to the episodes and personal experiences. However, long-term memory has unlimited storage capacity, since memories from a long time ago may be remembered in great detail. The decay rate of information is slow compared with the rapid decay from short-term memory [72] but Child's view [65] was that information does not decay but seems to be permanent in most circumstances.

If we present learning as the transfer of material from short-term memory to long-term memory, then both **prior learning** and **later learning** may cause interference: the more similar the interfering materials the more confusion there is between items in memory [66]. This transfer may require a good deal of attention [62].

A useful summary of the commonly accepted features of the three memory stores has been provided, in cases where the material to be remembered is verbal [70] (Table 1).

MEMORY PROCESSES

The human memory works analogously to a library [62]. It has an effective card-cataloging system, it appears to know what information has or has not been stored. By applying an appropriate search procedure, memory can **retrieve** and **recognize** any particular item. However, for this purpose, it uses some control processes such as **pattern recognition**, **rehearsal** and a **set of manipulative logic rules** such as induction and deduction which seem to play an important role in both storage of information (or learning) and problem solving.

When information enters into memory, it processes through several stages since the memory system provides the opportunity for processing inputs effectively and regulating the rate of flow of information [60]. The model (which is given in Table 2) illustrates the likely relationship between the memory stores based on **schema theory**. In this model, short-term memory is regarded as a working memory, possibly with "slave" systems like articulatory loops for rehearsing items temporarily and the emphasis has moved from a selective filter which automatically reduces processing overload to an active central

Feature	Sensory Registers	Short-term Store	Long-term Store
Entry of Information.	Preattentive.	Requires attention.	Rehearsal.
Maintenance of Information.	Not possible.	Continued attention. Rehearsal.	Repetition. Organization.
Format of information.	Literal copy of input	Phonemic. Probably visual. Possibly semantic.	Largely semantic. Some auditory and visual.
Capacity.	Large.	Small.	No known limit.
Information loss.	Decay.	Displacement Possibly decay.	Possibly no loss. Loss of accessibility or discriminability by interference.
Trace duration.	¼ - 2 Seconds.	Up to 30 seconds.	Minutes to years.
Retrieval.	Readout.	Probably automatic. Items in consciousness Temporal/ phonemic cues.	Retrieval cues. Possibly search process.

Table 1. Commonly Accepted Differences Between the Three Storages of Verbal Memory (from Craik and Lockhart [70]).

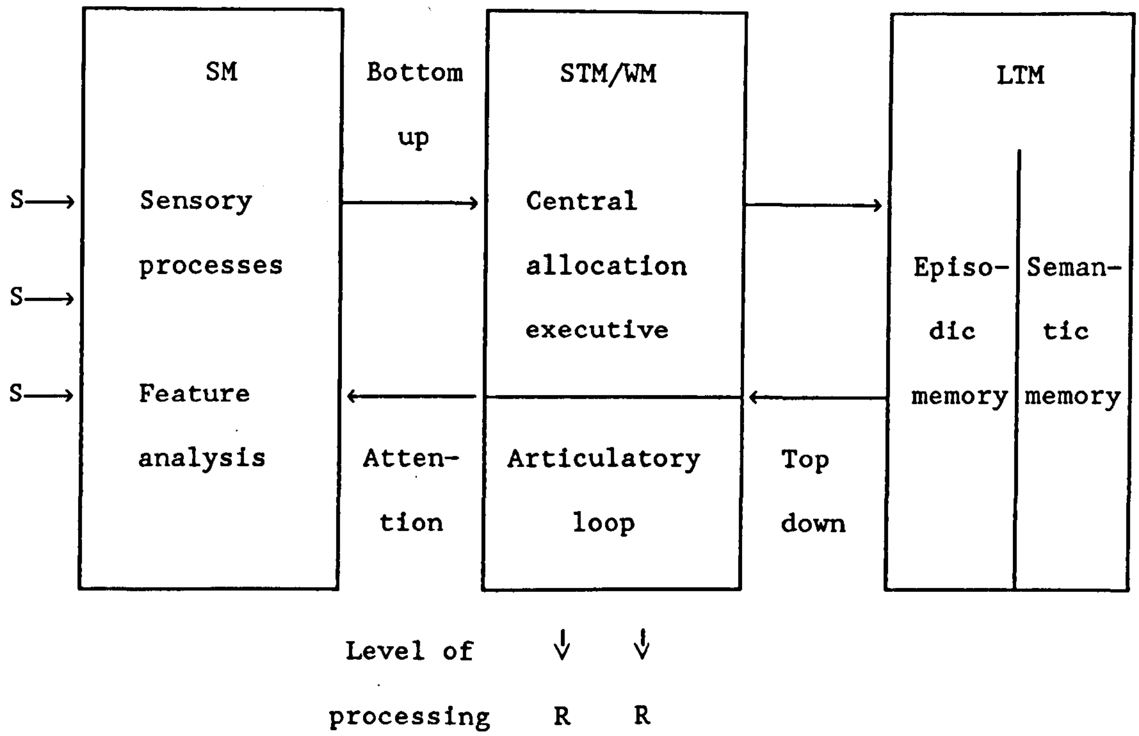

Table 2. Composite Model (from Roth and Frisby [66]).

processor which decides the level of processing input, accesses the knowledge from long-term memory and prepares items for output. The model has also been influenced by top-down schemes based on experiences and expectations stored in long-term memory [66].

The organization of knowledge in long-term memory and the working memory may be the keys for understanding how this model works.

ORGANIZATION OF KNOWLEDGE IN LONG-TERM MEMORY

Information processing theory deals with the capabilities of humans to understand, generalize and invent a richer conception of how people store and retrieve knowledge. It develops the notion of scheme for representing the organizational aspects of long-term memory. An example of this representation is a **semantic network** [62]. The semantic network consists of nodes, representing concepts, linked by lines which express the relationship between these concepts. These networks are known as concept maps. According to this view, knowledge is organized in interrelated chunks. It takes into account a number of mental capabilities and how people make inferences.

The human mind is an active, not a passive recorder of associations from outside. Therefore, knowledge becomes structured in meaningful ways rather than in a random collection of bits of information [3].

Semantic networks in the view of Stewart and Atkin, are models of how conceptual information might be stored in an individual's long-term memory. As such they can be used to give meaning to terms such as understanding or learning. An example of the knowledge structure for multiplication and division is illustrated by Figure 1. In this figure, each operation has a definition, an object and an outcome and both are the inverse of each other [3]. One can confirm that learning "more" usually results in better organized and linked knowledge rather than separate pieces of information [3]. This may lead to that major goal of mathematics instruction: to help students to acquire well structured knowledge about mathematics.

To sum up, information is better remembered if it is meaningful (in the Ausubel sense) and maybe non-meaningful material is more difficult to retrieve from long-term memory. The integration or association with other material may facilitate the retrieval. According to Ashlock et al. [8], the following procedures may aid the recall: "first letter cueing", the "Loci method", the "hook or peg system" and the "successive comparison technique", since all the above methods show superiority in helping remembering things over rote learning.

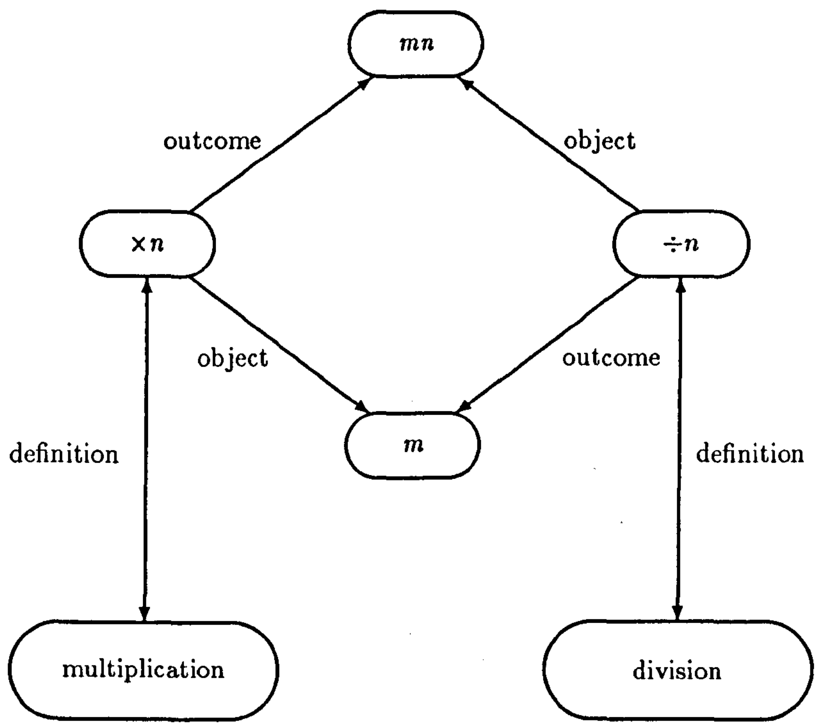
mn
outcome
object
$\times n$
$\div n$
object
outcome
m
definition
definition
multiplication
division

Figure 1

WORKING MEMORY

Baddeley (1981, [74]) and others have developed the theory of working memory which attempts to specify which components of the short-term memory system might be involved in various tasks. It was reported [75] that working memory is a workspace for holding information needed temporarily for the purpose of some other processing activities. This working space may be flexibly divided between data storage and data manipulation [67]. However, Johnstone (1984, [76]) noted the limitation of working memory in size and defined it as "that part of the brain where we hold information, work upon it, organize it, and shape it before storing it in long-term memory for further use."

The working memory has four components namely a central executive, an articulatory loop, a visuo-spatial scratch pad and a primary acoustic store [77]. It was noted [63] that the central executive is the most important of these components since it is involved in all tasks that require attention, and directs the operations of the other components. It has a limited capacity but it can process information in any sensory modality in a variety of different ways and store information over brief periods. The four listed components can be regarded as the "attentional system", the "inner voice", the "inner eye" and the "inner ear" respectively.

The role of working memory has been stressed [66] as selecting inputs, accessing long-term memory, planning strategies for solving problems and outputing appropriate responses. It is used [2] for executing most tasks and its properties depend upon the sort of the task being carred out. The limitation of working memory capacity may explain why all operations may not occur simultaneously without impairing its performance. Therefore, if two tasks are performed at the same time, they might interfere if they both use at least one of the common components. In other words as Baddeley suggested [78] learners are incapable of emitting two unrelated responses at the same time and that one response will occur before the second one which is almost invariably delayed.

The role of working memory in **mental arithmetic** has been examined [79]: the information in arithmetic operations, unless written down, is held in working memory rather than long-term memory. Therefore, errors are likely to increase as the amount of written information decreases. The following examples are ordered according to their difficulties:

345	345	...	...
+ 263	+ ...	+ 263	+ ...
____	____	____	____

It is also seems likely that the longer material is stored in working memory the more likely it is to be forgotten and that errors can be adequately explained by the loss of information in working memory.

To sum up, the working memory may provide a useful conceptual framework for exploring the nature of mental arithmetic.

LEVELS OF PROCESSING

The idea of depth processing was developed by Craik and Lockhart (1972, [70]). According to this theory [63] deep or semantic processing leads to better long-term memory storage than shallow or nonsemantic processing. It was noted that much of the field's evidence supports the basis of this theory particularly the affirmation that processing activities, at the time of learning, can have a major impact on subsequent retention.

The depth theory was interested in how processing activities affect long-term memory. Both working memory and levels of processing assume that people have flexibility in selecting what to attend to and how to learn; but this flexibility makes it more difficult to control what subjects are processing in psychological experiments [66]. However, the relationship between the two approaches could be explored in the level of processing in which different subjects are given different orienting tasks to perform.

OVERLOADING OF WORKING MEMORY

A list of possible reasons which may be considered as obstacles to grasping a concept has been suggested by Wilson [80]. One of these reasons is that "the concept is governed by so many rules that he (the learner) cannot keep them all in mind at once. By the time the teacher is explaining the last he has forgotten the first." This situation arises when many pieces of incoming data overload the capacity of working memory. Therefore, no processing of the data can take place since some are lost at the beginning.

Barber (1988, [60]) has the same idea: "if the information we are concerned with reaches the upper limits of our working space, an overloading in the capacity of working memory could occur. A loss in productivity may arise."

The overloading has been discussed in terms of the learner's performance. Johnstone [81] analysed the case when students encounter problems or learning situations of increasing load. He expects a good performance when the load is within the capacity of the working memory but, if the load exceeded their upper limit, the performance would drop suddenly. Factors such as perception, practical work and language may cause overload too. Much work has been done on this area (e.g. Johnstone [82], Letton [83] and Cassels & Johnstone [84]). The main findings of this work were that:

(i) a student who needs to recall, sequence and at the same time use more information is more likely to get the wrong answer (since the form of his perception of the total problem overloads his working memory);

(ii) a student's learning during practical work is more likely to be lost (since the noise of irrelevant information predominates over the signal);

(iii) a student is more likely to be overloaded when he deals with a negative question or unfamiliar vocabulary (since those require more working memory space and usually cause a loss in performance).

Moreover, overloading also may occur from a combination of difficulties arising from the nature of a subject matter, the method of teaching and the way of learning [76].

Minimizing the load is very important in facilitating the teaching and learning process. This may require cooperation with the limitation of working memory capacity. The following may help to achieve this purpose:

(1) the information content should be kept low;

(2) redundant and irrelevant information should kept out;

(3) the employment of language should kept simple and familiar;

(4) the transmission of knowledge should ensure understanding since rote learning (which easily occurs in this operation) may not guarantee this purpose;

(5) when the information content is necessarily high, due to the nature of material, then providing some "rule of thumb" may help the material to be chunked; in short, the remedy of overload may lie in chunking and grouping information.

CHUNKING

It is clear now that short-term memory can carry only a very small number of items since its storage consists of a finite number of "boxes" or "slots", each of which can store one item of information and the number of slots increases with age [8]. However, the term chunk (i.e. box or slot) has been used firstly by Miller in his very influential paper (1956, [71]) to indicate a word, letter or digit which describes a familiar item or unit. Johnstone and Kellet [85] defined a chunk as "what the observer perceives as a unit, for instance a word, a letter or a digit." Reed [72] considered a chunk as a group of items that is stored as a unit in long-term memory.

According to Miller, the ability to receive, process and remember information depends on the span of absolute judgment and the span of immediate memory which are likely to be limited by the amount of information and the number of items respectively. He made a distinction between bit and chunk but each has a constant number for absolute judgment and immediate memory. He also noted the independence between span of immediate memory and the number of bits per chunk. Although the memory span is a fixed number of chunks, increasing the number of bits of information per chunk is still possible. This leads to building large chunks, each containing more information than before. Simon [86] stated that "the change with age in the digit span of the human memory is due to the growth of encoded strings in the human's chunks."

An example of chunking is the telephone number of Glasgow University: 041-339 8855. This number can be regarded as three units each unit is packed into three or four items and each package makes a chunk. Therefore, it is much easier to process it as three chunks rather than as ten separate digits.

Grouping or chunking certainly reduces the memory load and this underlies much, if not all learning behaviour. Overload occurs when the number of separate bits of information overcharges the capacity of short-term memory.

A major determinant of individual difference in memory is how effectively people can group material into familiar chunks. It was reported [87] that pupils with a low level of conceptual understanding are disadvantaged since they chunk inefficiently, treat redundant information as necessary and use inefficient or arbitrary strategies in high information contexts.

It is known that the span depends on the familiarity of the items. Therefore, items are stored in chunks where the number of items in a chunk is dependent on the meaningfulness, or relatedness, of the items [8]. If we consider 7 ± 2 chunks as the capacity of short-term memory then the chess master's 7 chunks for example, are much richer than those of the novice since the former has larger units than the latter even though they manipulate the same number of units [2]. The interaction between information content, the state of conceptual development and the perceived level of difficulty has been suggested by Johnstone [85] who stated that:

(1) the number of units represented by the information will depend upon the conceptual understanding;

(2) the larger the number of chunks, the more difficult the material will seem to be and the poorer will be the results;

(3) if the chunk capacity is exceeded, two possible results will appear:

   (a) the pupil will extract no useful information if he tackles the problem as a whole; or

   (b) if he has some memory saving strategy which allows for sequential treatment, he may succeed;

(4) conceptual understanding leads to an efficient organized and converging strategy.

To sum up, the psychological reality of the chunk as Simon noted [86] has been fairly well demonstrated and the chunk capacity of short-term memory has been shown to be in the range of five to seven.

NEO-PIAGETIAN THEORY

The weaknesses of Piaget's theory led to a number of attempts to relate its stages of cognitive development to the development of short-term memory capacity by describing tasks at the different stages in terms of concepts or schemes that have to be considered simultaneously. As an example of these attempts, Mclaughlin [88] proposed that the number of concepts that need to be coordinated in preoperational, concrete and formal operational tasks is 2^1, 2^2 and 2^3 (or 2, 4 and 8) respectively. Therefore a child would need a working memory capacity equal to the number of concepts to cope with tasks at a given stage. However, Halford [89] agreed with capacity 2 and 4 but he suggested 6 rather than 8 for formal tasks.

The well-known alternative approach is **Neo-Piagetian Theory** which was proposed by Pascual-Leone and Smith (1969, [90]). Its goal is to make Piaget's theory functional not just structural as it is. This gives the theory the strength of predicting performance of students of a given ages since the emphasis is more on the child's processing capacity rather than storage capacity [8] and the description of mechanisms by which knowledge is acquired and put to use [5].

The basic notion of the theory is a **scheme** or unit of thought which represents experience and produces behaviour. According to their function, schemes can be classified into **figurative**, **operative** and **executive**. It was reported [91] that:

(i) figurative schemes (or chunks) are the internal representations of items of information which are familiar to a subject;

(ii) operative schemes (or transformation or primitive information processing) are the internal representations of functions or rules which can be applied to one set of figurative schemes, in order to generate a new set;

(iii) executive schemes (or plans or executive programmes) are the internal representations of procedures which are applied to a problem in order to obtain a particular objective; they are responsible for determining what figurative and operative schemes are to be activated in any particular situation.

It was noted that all these schemes are active, functional units and have a releasing and an effecting response. But, because they are internal and subjective, they are difficult to measure.

According to Case [92], the child is born with an innate repertoire of sensory motor schemes and then he applies and modifies his basic repertoire of schemes during everyday interaction with the world. Modification and combination of old schemes are very efficient for acquiring new ones.

The process of thought during working on a problem may be characterized by activating some executive schemes which direct the activation of a sequence of figurative and operative schemes. This sequence consists of separate mental steps. However, because mental efforts which are required to rehearse any of these schemes are limited, the number of schemes used in any one mental step is limited too. Finally the executive schemes direct the response when they reach their maximum. It was noted that the nature of the problem, the perceptual field, the experience and the emotional reaction to the situation all affect these schemes and hence the response. Therefore, the success in problem solving depends on the following:

(1) the repertoire of schemes which increase in complexity and accuracy with learning and maturation;

(2) the maximum number of discrete schemes which an individual can activate simultaneously through an attending act;

(3) the tendency to use the full mental space that is available to a subject;

(4) the relative weight which is given to cues from the perceptual field rather than from task instuctions, in selecting an executive scheme.

The mental capacity may vary with age and biological and maturation factors. Its size is assumed to increase linearly wih age according to the following scale:

Age	Developmental stages	Maximum value of mental capacity
3 - 4	Early preoperations	$e + 1$
5 - 6	Late preoperations	$e + 2$
7 - 8	Early concrete operations	$e + 3$
9 - 10	Middle concrete operations	$e + 4$
11 - 12	Late concrete - early formal operations	$e + 5$
13 - 14	Middle formal operations	$e + 6$
15 - 16	Later formal operations	$e + 7$

Table 3. The relationship between age, Piagetian development level and processing capacity.

In this scale, the constant *e* refers to the space required by the executive schemes, while the numeral refers to the maximum number of additional operative or figurative schemes which can be activated under the direction of this executive.

The last two factors are highly correlated and form what researchers called cognitive style or field dependence-independence (e.g. Witkins et al. [93], Pascual-Leone & Smith [90]). It was noted that field dependent subjects may be characterized by being habitually low mental processors and highly influenced by the perceptual field being easily distracted by irrelevancies, whereas field independent subjects are assumed to be the opposite. One therefore may expect a lower or higher success rate depending upon a subject's ability to overcome the influence of irrelevant information in a surrounding field or to separate an item from its context.

The mental capacity (or central computing space *M*) is a very important factor and Pascual-Leone [27] attempts to explain cognitive growth by its size. He distinguished between the structure *M* and its functional use. The former is the maximum available capacity, whereas the latter is the amount of space used at any point in cognitive activity. Therefore, functional *M* may vary from zero to maximum capacity and it is moderated by a several factors such as the degree of familiarity and field dependence-independence which could influence the performance level of the subject. This opinion of familiarity is modified since, at the beginning, the belief was [91] that "the familiarity with the task does not have a large effect on performance" and linguistic competence is a consequence of thought as opposed to the ability to think being dependent upon linguistic development" [90]. This view is similar to that of Piaget which contrasts with other evidence in the research field (e.g. Cassels & Johnstone [84]).

According to Niaz (1987, [94]), the development of formal operational thinking, *M*-space for information processing , the ability to disembed relevant information and previous experience are all factors which affect performance. It was reported [91] that performance will depend on the content of the child's repertoire independently of the magnitude of the child's *M*-space. But misleading schemes in a subject's repertoire must be extinguished [95]. However, in his experiment of decoding-encoding, Pascual-Leone [90] predicts that mental age, form of representation and the task complexity all influence the attainment in problem solving tasks.

The task difficulty (or the demand of a task or information processing load) is another factor which plays an important role in determining the success on a task. It is defined by Scardamalia [96] as the maximum number of schemes that the subject must activate simultaneously, through an attentional process, in the course of executing a task. It was noted that the same task may have a different demand for different subjects depending on the schemes they coordinate and the manner in which they chunk information presented to them.

Scardamalia noted that the quality of a solution depends on the information load which produces failure or success or break down and that the logical capabilities of subjects can be grossly misjuged if we do not present a task in its lowest possible load corresponding to its logical structure. She explained the "Horizontal Decalage" (the phenomenon of passing certain tasks and failing others with the same logical structure) in terms of information processing load which increases with the logical complexity of tasks, but may also vary within tasks of the same logical structures (e.g. mental multiplication of numbers of different sizes). This opinion is similar to that of Niaz [97].

The practical value for determining the difficulty of a task, according to Ashlock [8] is limited for two reasons: first, the analysis required to determine the schemes relevant to a particular task can be extremely complicated (e.g. Pascual-Leone [90]) and ambiguous (e.g. Lawson [98]). Secondly, other factors such as the familiarity of a cue for a given individual, the salience of a cue for a given type of task and field dependence-independence can affect task difficulty. It was noted that the Neo-Piagetian theory can only be fully tested in situations where the subject's mental strategies and task analysis are unambiguous [91].

Case [99] noted disagreement between developmental psychologists not just on how to compute the quantitative load that a strategy places on a child's working memory, but also on whether the measured growth in children's working memory has a functional or structural basis. However, the Neo-Piagetian theory needs to be heeded for its educational implications since it underlines that the complexity of a task and the limitation of a processing capacity can have a crucial effect when we deal with teaching material and problem solving tasks.

The contribution of Case on neo-Piagetian theory appeared in his various writings (e.g. [100], [101], [102] and [103]) which indicate very positive work on areas such as:

(a) the success not only depends on a subject's capacity but also on the demand of a task;

(b) instruction can affect task success when the load of a task exceeds the subject's capacity;

(c) the prediction of success is dependent upon careful task and strategy analysis;

(d) across task validation.

Case has interpreted the subject's capacity as a necessary but not sufficient condition for success. For him a child would succeed in a task when it is administered as a transfer item or when he is able to reduce the number of schemes to be coordinated to hypothesized capacity. He defined the success as a function of mental strategy, the demand of a task and the mental capacity. Therefore, by using these parameters, the qualitative characteristics of Piagetian stages can be accounted for in terms of quantitative ones.

The complexity of cognition and difficulty of teaching are features of both cognitive development and classroom tasks. Therefore, the source of difficulties may be one of inappropriate strategies, or the instructions overloading the working memory or insufficient familiarity with the basic operations. To overcome these difficulties, Case devised a method of instructional design based on the following three stages which are summarised by Macnab and Cummine [61]:

(1) find out how children might try to perform the task if they were not told how to do it;

(2) motivate the learning of the preferred method to be taught by making clear the limitation of pupil's naive ideas;

(3) in designing learning hierarchies, keep working memory requirement as low as possible.

In practice, there are difficulties in applying the first two stages since they tend to deal with mistakes which the pupils may never make or hypothetical reasons for mistakes which are not in fact the real reasons. This led Gagné [104] to argue against Case's analysis when he suggested that "teachers would best ignore the incorrect performance and set about as directly as possible teaching the rules for correct ones." But teaching only correct methods may lose the opportunity of discussing the areas of misconception that experience suggests pupils may have.

The third stage is very important in learning and problem solving since it introduces the notion of working memory. As a very simple example of keeping the load in working memory as low as possible, finding the value of x in

$$34 + x = 51$$

by means of an add-on procedure may require lower loading than realising that the answer is given by 51 - 34 and performing this by a formal algorithm. But a really effective algorithm can substantially reduce working memory requirements too. The formula

$$\tfrac{1}{2}n(n + 1)$$

for the sum of the first n whole numbers is an example.

According to Case, the neo-Piagetian theory has the power to predict the interaction between instruction and development, provided the possible mental strategies can be specified and assessed. Success can be predicted provided that:

(i) the strategy can be taught;

(ii) the minimum demand of a task can be assessed using a well defined procedure;

(iii) one can analyse the task for misleading cues and overlearned responses.

To sum up, the neo-Piagetian theory is relevant to the competencies of Piaget by its formal aspect and has the power of prediction by its functional aspect. It seems to bridge the gap between development and learning theories since it provides both perspective within the one framework and it make precise predictions possible [91].

A PREDICTIVE MODEL BASED ON INFORMATION PROCESSING THEORY

We have already discussed the information processing approach which in general reflects the way in which the memory system encodes, stores and retrieves information. An example of this approach is the model of Johnstone and El-Banna (1987, [5]) which relates students' performance to the amount of information to be processed in a learning or problem solving situation [105]. This model has benefitted from the work of Duncan and Johnstone [106], Johnstone and Kellett [85], and Johnstone and Wham [107]. It re-examines these earlier works and adds to them some very effective factors which control a pupil's ability to interpret and handle questions. Such factors are:

(a) the need to reconstruct the meaning for one's self;

(b) the limitation on the size of the working space;

(c) the noise which swamps the signal;

(d) the tendency to be distracted by irrelevant information [108].

It was also noted that knowledge has to be reconstructed as it passes from one person to another and what we already know and understand controls how we interpret, process and even store information. The part of the brain [109] where conscious processing takes place is of very limited capacity. This shared area permits one to hold ideas and think about them in terms of encoding, ordering, application of rules and pattern seeking. The new ideas coming in can displace ideas already there unless efficient grouping and chunking take place. The constitution of a chunk is controlled by previous knowledge, experience and acquired skills.

According to Johnstone, a common factor of working memory overload seemed to appear in all areas of science (and also mathematics) which pupils perceived to be difficult. Such load may be found in the laboratory in terms of noise and signal, language in terms of familiarity and negative forms, material in terms of density of information and the subject itself in terms of its nature. In addition to these sources of overload, teachers may unwittingly make a wrong estimate of their pupils [81].

The results (Johnstone [108]) when a working memory overload occurs are:

(i) the impossibility of giving any answer or;

(ii) breaking down the task and dealing with a small portion at a time or;

(iii) developing a strategy which permits grouping and chunking of information.

He explained the distraction by irrelevant information in the sense of working memory space: pupils of low working space do not have enough space to take any irrelevant material, while those of large working space may have enough excess space to perform successfully despite irrelevancy.

The model deals [2] with three main parameters (see figure 2):

(1) the working memory capacity (X) that the learner has;

(2) the mental strategy (Y) that the learner uses to solve a task;

(3) the demand (Z) that the problem puts on the learner's mental capacity.

The working memory capacity increases with age at the rate of one unit every two years up to maturity. Its size 7 ± 2 has been affirmed by Miller [71] as the capacity to hold information without

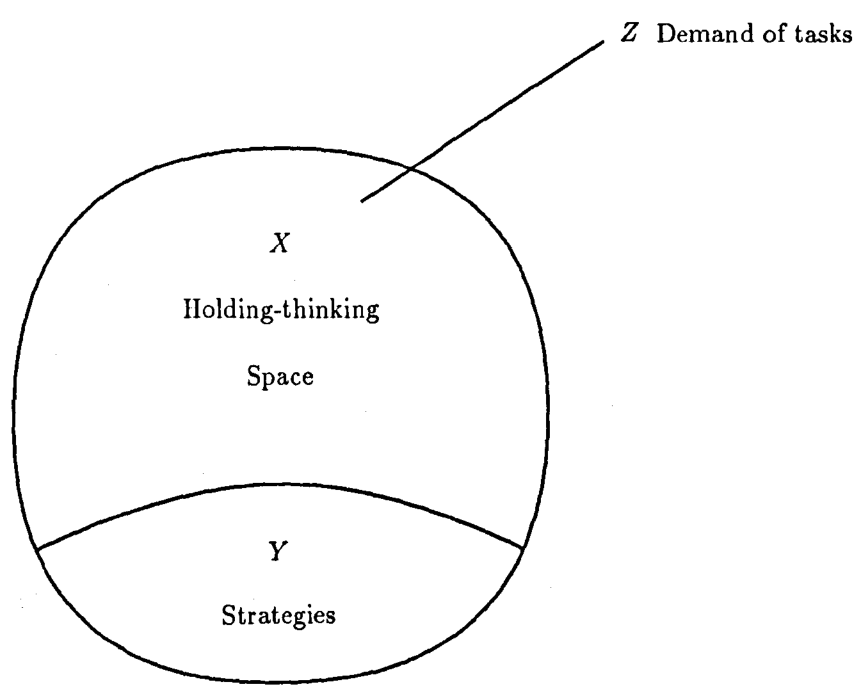

Figure 2

manipulating it. Whereas Johnstone [2] suggested the size 6 ± 2 to allow some space for this manipulation.

Strategy is an essential factor in teaching and learning. It can be developed and improved. This leads to reduction of the demand of a task and hence to the improvement of performance. It has been found that [110] the higher processing capacity pupils are able to choose and apply an appropriate strategy rather than a trial-and-error method.

The demand of a task according to Johnstone and El-Banna [105] is the maximum number of thought steps and processes which had to be activated by the least able, but ultimately successful candidate in the light of what had been taught. As an operational method for assessing it, they analysed "numerical questions" in terms of what had been recalled, transformed, deduced and concluded. Therefore, the following three factors are those which (in an interaction situation) could give the maximum demand:

(i) the information in the question which has to be processed;

(ii) the information which has to be recalled to add to the given information;

(iii) the processes which have to be activated to deal with the information, processes such as deduction, transformation and calculation.

An illustration of the first factor can be given in terms of language, negative forms, the way of the data is arranged, the existence of much irrelevant information and unnecessary data, etc. The second factor may include the recall of formulae, definitions and theories etc. The third factor may require the insight to see and deal with appropriate operations.

The above method - according to Johnstone - may be just a rough estimate since the maximum demand on working space is likely to take place before any sequencing is achieved, but it permits us to arrive at a relative indication of demand. The interaction between the model's parameters lead to success or failure in learning, teaching and problem solving situations. The model's hypotheses are [105] that:

A **necessary**, but **not sufficient**, condition for a student to be successful in a question is that the demand of the question should not exceed the working memory capacity of the student. If this capacity is exceeded, the student's performance will fall unless he has some strategy which enables him to structure the question and to bring it within his capacity.

A **negative correlation** will be found between the percentage of successful attempts at a question (facility value FV) and the demand of the question (*Z*). The curve which represents this correlation is S-shaped indicating a high and a low plateau with a rapid drop between them (see figure 3).

These characteristics can be interpreted in the light of working memory space which is capable of processing about 5-7 pieces of information:

(a) the **upper plateau** must consist of questions which were within the working memory capacity of all, but carelessness, forgetfulness or lack of interest could depress the achievement from 100%;

(b) the **hole in the middle** (Duncan and Johnstone could not interpret it at the time, 1974) appears between five and six on the demand axis. It shows an overload in processing

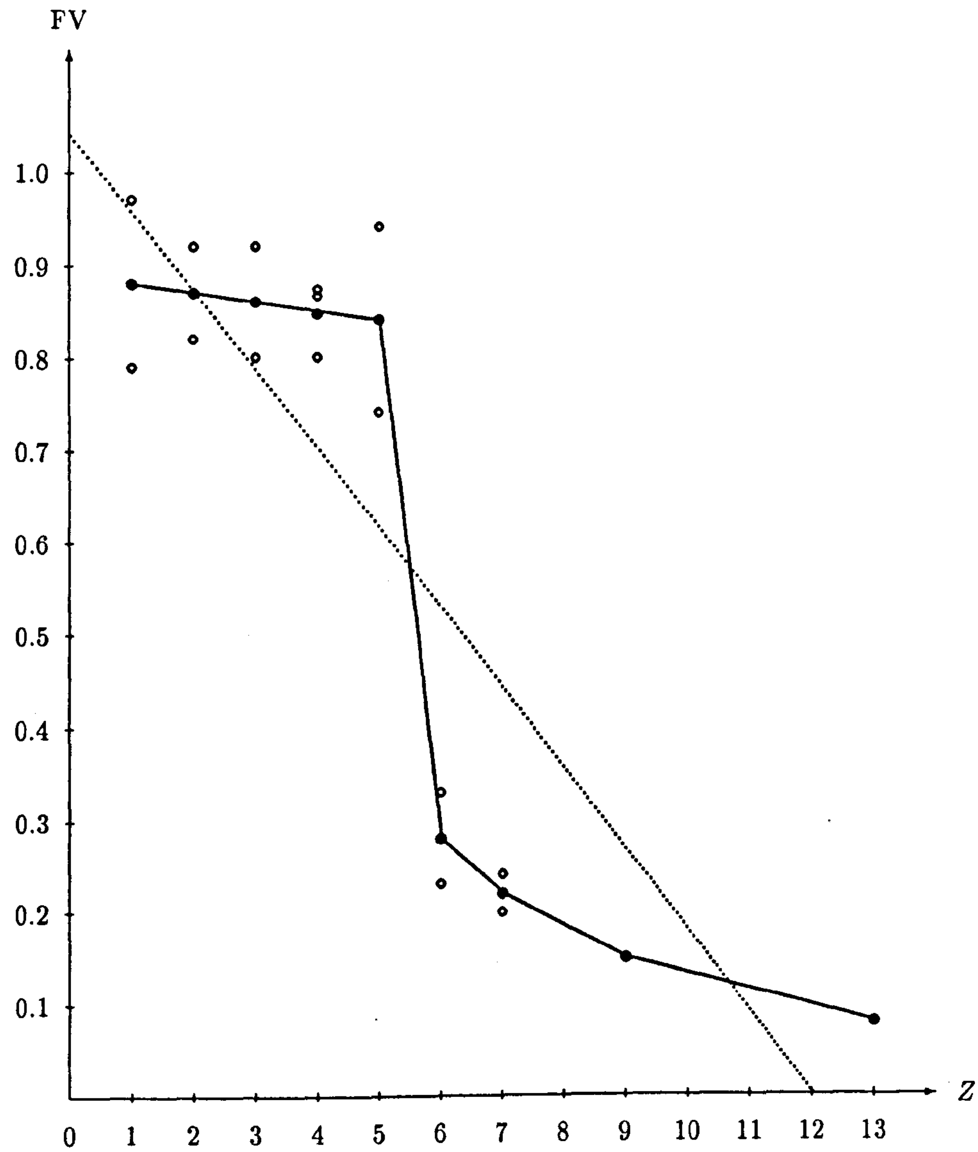

Figure 3

capacity because of the nature of the subject or the way in which it was being taught and learned or some combination of these;

(c) the lower plateau represented the questions which had now gone beyond the most pupils, but a few may continue to function successfully according to their strategies which enable them to overcome their capacity limitations.

To sum up, the model brings and puts together many efficient ideas [109]:

(i) the idea of "capacity increasing with age and the skills for using it" is Piagetian in nature;

(ii) the idea of "strategy includes pre-learned concepts which enable incoming ideas (in the sense of its demands) to be processed and meaningfully learned" is an Ausubelian one;

(iii) the idea of "pupils own science" is part of the prelearning which affects formal learning and brings the alternative framework ideas into strategies.

The interaction of these ideas lead to the predictive power of this model which gives an opportunity to raise and test the above hypotheses in many subjects areas (i.e. science and mathematical subjects).

The educational applications of this model ([5], [109] and [108]) are certainly varied from the content structure of material to be learned, across teaching methods to assessment. Some of these implications are briefly summarised in the following assertions:

(1) the traditional presentation of scientific facts and concepts must be re-examined in the light of the demand of a task and students' capacity;

(2) sequencing and organizing knowledge, dealing with a bit of information at a time and using familiar language provide a great help to the student;

(3) a student must be helped to develop his own strategies and given the opportunity to practise strategies in terms of breaking down a task into its parts, dealing with high information loads and separating relevent from irrelevent information;

(4) the high demand of a question should be reconsidered since it tests both capacity and strategy (care should be given to the amount of information to be manipulated and the question's language in terms of the degree of familiarity and negative forms).

Finally, it would be unjustified to suggest [105] that the interpretation of all problems of learning, teaching and testing can be made by this model, but the suggestion of a mechanism for some problems which do exist and a mechanism by which they might be overcome is certainly possible and obtainable.

CHAPTER THREE

Application of the Holding-Thinking Space Model in Tertiary Level Mathematics

INTRODUCTION

In the previous chapter, an information processing model was described and discussed which considered a shared holding-thinking space (or working memory space, which an individual might possess). The capacity of this is the maximum number of pieces of information which can be held and operated upon at any given time during the working on a task.

It has been found that a subject's performance depends on his working memory capacity, strategies which he may employ to solve a task and the task's complexity. He will fail to solve the task if its load exceeds his working memory capacity, unless he uses an appropriate strategy which may help him to succeed.

In order to test the model's hypotheses, research was carried out upon the work of students in the two first-year mathematics classes at Glasgow University. The work began in October 1987 and the starting point was to accept that the capacity of the students would be about 6. The demand of examination questions according to the number of thought steps was determined and the students' performance on these questions was analysed. In short, the subjects' progress in mathematics was traced for two university class examinations in 1987-88, and comparison between performance, working memory capacity and the demand was made in both "A" class and "B" class samples.

The procedure of assessing the complexity of questions [105] was to choose at random 25 scripts where students had scored full marks in a given question and to examine their routes in terms of what had been recalled, transformed, deduced and concluded. The number of thought steps in the longest route was taken to be the demand. This process was repeated for each question.

As a result of the above work, comparison between performance, capacity and the demand has been made in the light of the theoretical predictive model of Johnstone and El-Banna (see Figure 4) in which the subject of average capacity six will perform well if the demand of a question is equal to or less than six, but his performance will drop down rapidly if the demand exceeds six unless he has some space saving strategy.

The "A" class sample

The sample contained 163 students.

The first term examination (November 1987)

The exam was composed of sixteen items (parts of questions), but just eleven of them were considered. The other five were omitted since four of them were theoretical and one contained an error. The theoretical items were omitted because it is possible to answer them using only recall. The items were then grouped according to their demand into six sets of complexity 2, 6, 7, 8, 9 and 10. The analysis of the demand of the items was obtained from students' scripts, and the agreement about the total number of thought steps was made by two teachers. The following examples illustrate such an analysis.

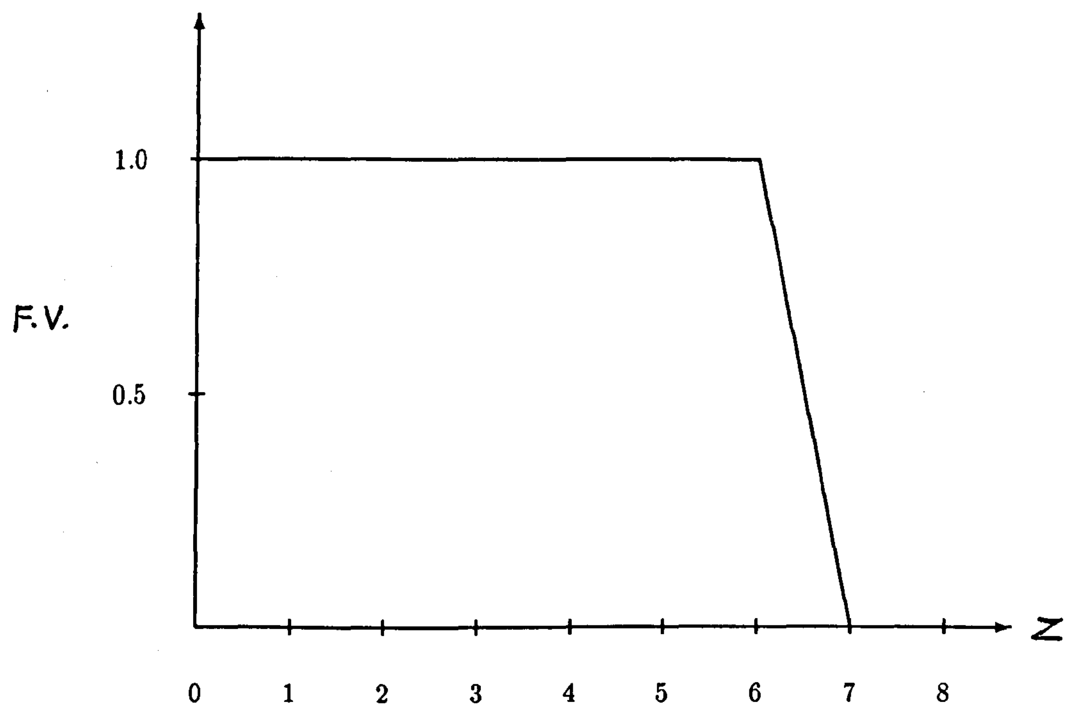

Figure 4

Example 1

Simplify $\dfrac{1}{\sqrt{(1 - \frac{1}{1 + x^2})}} \cdot \dfrac{x^2}{1 + x^2}$.

The question could be analysed into the following steps:

step 1 simplify the expression $1 - \frac{1}{1 + x^2}$,

step 2 recall that $\sqrt{\frac{a}{b}} = \frac{\sqrt{a}}{\sqrt{b}}$,

step 3 use $\dfrac{1}{\frac{\sqrt{a}}{\sqrt{b}}} = \dfrac{\sqrt{b}}{\sqrt{a}}$,

step 4 recall that $\sqrt{a} = a^{\frac{1}{2}}$,

step 5 cancel $(1 + x^2)^{\frac{1}{2}}$,

step 6 recall that $\sqrt{a^2} = |a|$,

step 7 cancel $|a|$.

Therefore the *Z*-demand of this question is equal to 7 (thought steps).

Example 2

Prove that, if $a > b > -\frac{4}{3}$, then

$$\frac{a + 1}{3a + 4} > \frac{b + 1}{3b + 4}.$$

The thought steps of this question are:

step 1 consider the difference between the two given fractions,

step 2 unify the fractions over a common denominator,

step 3 multiply out both products in the numerator,

step 4 multiply the terms in second product by -1,

step 5 simplify the numerator,

step 6 determine the sign of the fraction, step 7 deduce the result.

Therefore the Z-demand of this question is equal to 7 (thought steps).

Example 3

Show that, provided $\theta \neq k\pi \pm \frac{\pi}{3}$ $(k \in Z)$,

$$\sin \theta = \frac{\sin 3\theta}{2 \cos 2\theta + 1}.$$

Deduce the values of $\sin \frac{\pi}{12}$ and $\cos \frac{\pi}{12}$.

This example contains three parts: the justification of the equality, and the determination of values of $\sin \frac{\pi}{12}$ and $\cos \frac{\pi}{12}$. Both the second and third part could be solved without the solution of the first part, but finding the value of $\sin \frac{\pi}{12}$ (or $\cos \frac{\pi}{12}$) helps to find other value. Therefore, it was decided to consider the whole question as two items (the equality and finding the values).

(a) The thought steps of the first item are:

step 1 expand $\sin 3\theta$,

step 2 expand $\cos 2\theta$,

step 3 expand $\sin 2\theta$,

step 4 factorise the numerator,

step 5 use $\cos^2 \theta = 1 - \sin^2 \theta$,

step 6 simplify the expression.

Therefore, the Z-demand of this item is equal to 6 (thought steps).

(b) The thought steps of the second item are:

step 1 substitute $\frac{\pi}{12}$ for θ,

step 2 substitute the values of $\sin \frac{\pi}{4}$ and $\cos \frac{\pi}{6}$,

step 3 simplify the expression,

step 4 recall that $\sin \frac{\pi}{6} = 2 \sin \frac{\pi}{12} \cos \frac{\pi}{12}$,

step 5 substitute the values of $\sin \frac{\pi}{6}$ and $\sin \frac{\pi}{12}$,

step 6 obtain an expression for $\cos \frac{\pi}{12}$,

step 7 replace the fraction in the denominator by its inverse in the numerator,

step 8 simplify the expression.

Therefore, the Z-demand of this item is equal to 8 (thought steps).

The complete November exam and analysis of other items can be found in Appendix 1.

It is important to realize that the maximum demand may occur when the student first reads an item and tries to take in all that he is being asked to do. During this period, he is turning the question over to see a way to start and how to proceed. This is the first stage, the second one could be that he sees this and sequences the problem. The demand must now drop and he needs only to manipulate a few ideas at a time at each step.

The method adopted here cannot give the absolute demand, only the relative demand as set out in students' working which is a reflection of the second stage and, by implication, is related to, but not necessary equal to, the absolute demand of stage one. The best we can hope to do is to place the questions in demand order, but stage one may involve factors not shown in students' subsequent working.

The facility value (the proportion of the sample solving a question correctly) for each item was calculated and tabulated (Table 4). In this and many other tables, facility values are shown as percentages. The table contains the number of subjects who succeeded in each item (*N*), the facility value (FV) (as a percentage) and the demand (*Z*). Table 5 shows the frequency of items which have the same demand and their average facility value. The results are illustrated in Figure 5.

As we expected, a strong **negative correlation** between the two variables was obtained ($p < 0.01$) and a **sharp drop** of the curve comes after 6 on the *Z*-axis (the number of thought steps). This led us to conclude that the students performed well when the demand of questions is equal to or less than six, but their performance drops down rapidly when the demand exceeds their capacity. It is also noted that the curve neither reached 1 nor dropped to 0. This may be explained by the existence of other factors (such as the degree of familiarity, the structure of questions the strategy which was employed etc.). These factors may play a crucial role in determining the way in which individuals deal with a task.

It was noted - as theoretical background - that the curve contains a high and a low plateau with a rapid drop between them. As we expected, there is a highly significant difference ($p < 0.01$) between the mean score of FVs of items with $Z < 6$ and that of items with $Z > 6$. In general, the upper plateau consists of items which were within the working memory capacity, whereas the lower plateau contains items which were beyond most of the students.

item*	1				2			3			4
	i	ii	iii		i	ii	iii	ii	iii		iv
			1	2					1	2	
N	2	59	68	26	117	24	2	8	22	89	18
FV (%)	1	36	42	16	72	15	1	5	14	55	11
Z	7	7	6	8	6	9	10	10	8	2	9

Table 4. The FV for each item of the 1st term exam of the "A" class. (*): five items were omitted, four of them are theoretical and one has an error.

Frq	1	2	2	2	2	2
FV (%)	55	57	19	15	13	3
Z	2	6	7	8	9	10

Table 5. The average FVs of items which have the same demand.

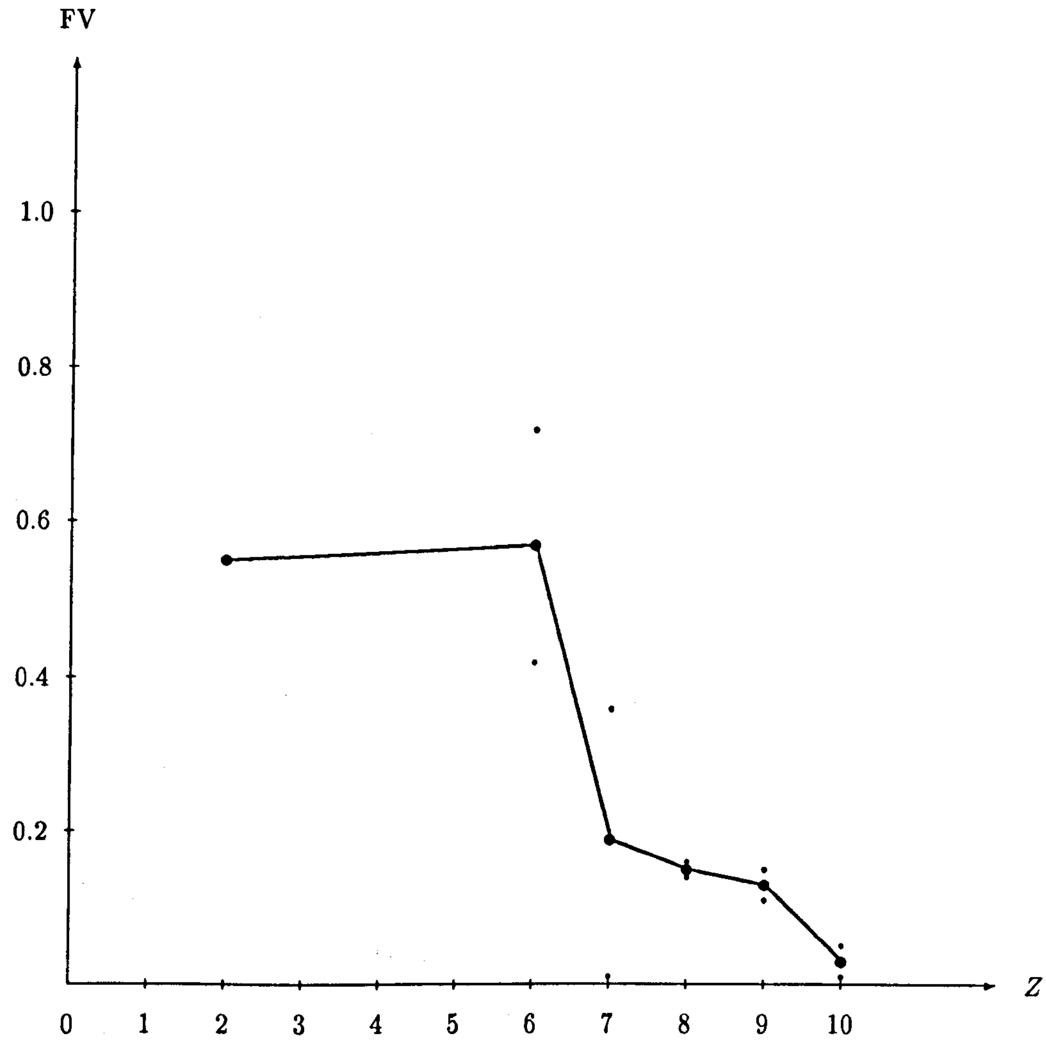

Figure 5

Using mean score to interpret data

The facility value is the proportion of students solving a question correctly, while the mean score is the average of their marks on this question scaled to produce a maximum possible mark of 10. Even though the holding-thinking space model predicts properties of the facility value rather than mean score, it is reasonable to interpret the data by considering the second factor. Table 6 gives mean score and standard deviation for each item, while Table 7 shows the frequency of items which have the same demand and their average mean scores. These are illustrated in Figure 6.

The pattern in general, looks like that of facility value except at the end where the performance of subjects on item with $Z = 10$ is higher than of that with $Z = 9$. This may due to the nature of mean score: students may have more of a chance to collect partial marks from the individual steps of the question with $Z = 10$ than that with $Z = 9$.

The influence of question's demand on subjects' performance has been tested. The question's demand in general, has an affect on the students' performance: the mean of the mean scores of items with $Z < 6$ is greater ($p < 0.02$) than that of mean scores of items with $Z > 6$. Note that a negative correlation between mean score and the demand was obtained ($p < 0.05$) as well as plateaus as before.

item	1				2			3			4
	i	ii	iii		i	ii	iii	ii	iii		iv
			1	2					1	2	
mean	3.6	5.6	5.3	4.0	8.3	3.9	4.3	4.7	4.6	6.5	3.8
SD	3.6	4.0	4.5	3.8	2.8	3.9	2.4	2.8	3.3	4.0	3.2
Z	7	7	6	8	6	9	10	10	8	2	9

Table 6. The mean score and standard deviation for each item of the 1st term exam of the "A" class.

Frq	1	2	2	2	2	2
mean SD	6.5 4.0	6.8 3.7	4.6 3.8	4.3 3.5	3.8 3.8	4.5 2.6
Z	2	6	7	8	9	10

Table 7. The averages of mean scores of items which have the same demand.

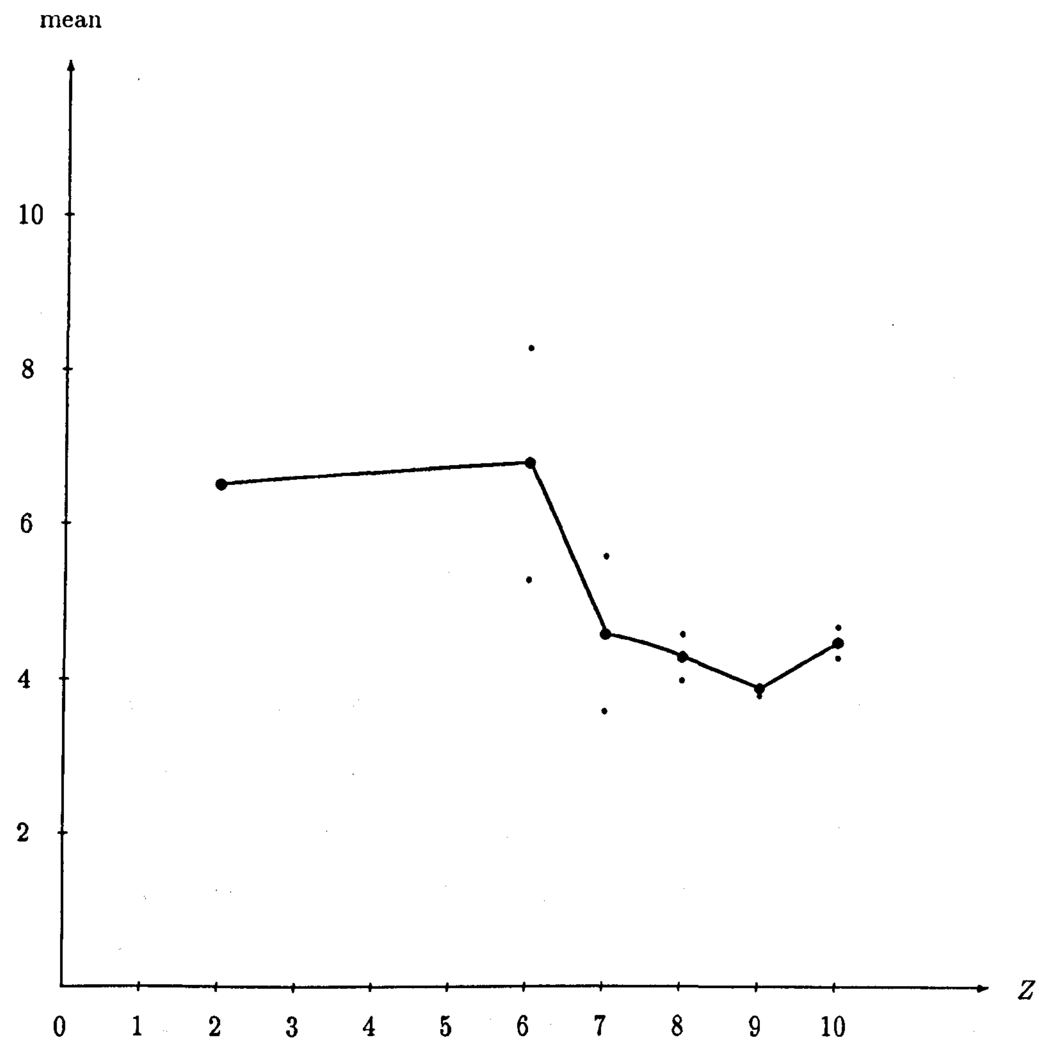

Figure 6

The second term examination (April 1988)

The same procedure used for the November exam was adopted here. In order to follow the students' achievement, the same sample was kept but nine students were absent from the exam. The size of the sample now is 154. The exam was made up of fourteen items but one of them was theoretical. The thirteen items were grouped into five sets of complexity 4, 7, 8, 9, and 17. The complete April exam and the analysis of the demand of the items can be found in Appendix 2.

The facility value for each item was calculated. Table 8 shows the results, while Table 9 gives the average facility values of items which have the same demand. The relationship between the two variables (facility value and the demand of questions) is shown in Figure 7.

Because there were no questions with $Z = 5$ or $Z = 6$, the students' performance dropped down from the question with $Z = 4$. The results show that the phenomenon of "no difference" in subjects' performance on questions of different demand (such as questions with $Z = 4$, $Z = 8$ and $Z = 9$) has emerged to indicate the influence of other factors in the achievement of students. An attempt was made to find out why the average FV of the questions with $Z = 4$ was like the average of the FVs of those with $Z = 8$ and $Z = 9$, and why the performance on the question with $Z = 7$ was weaker than the average performances on those with $Z = 8$ and $Z = 9$.

There were two items with $Z = 4$, viz 1(i) and 6(ii.1). The difficulty comes from the second one since it characterized by:

(i) the need to recall the formula for the roots of a quadratic equation;

(ii) the need to apply the condition on the discriminant for

Q.	1			2	3		4		5		6		
	i	ii	iii		ii	iii	i	ii	a	b	i	ii	
												1	2
N	72	43	37	17	29	62	61	49	80	71	40	35	33
FV (%)	47	28	24	11	19	40	40	32	52	46	26	23	21
Z	4	8	7	17	9	8	8	9	8	9	8	4	8

Table 8. The FV for each item of the 2nd term exam of the "A" class.

Frq	2	1	6	3	1
FV (%)	35	24	35	32	11
Z	4	7	8	9	17

Table 9. The average FVs of items which have the same demand.

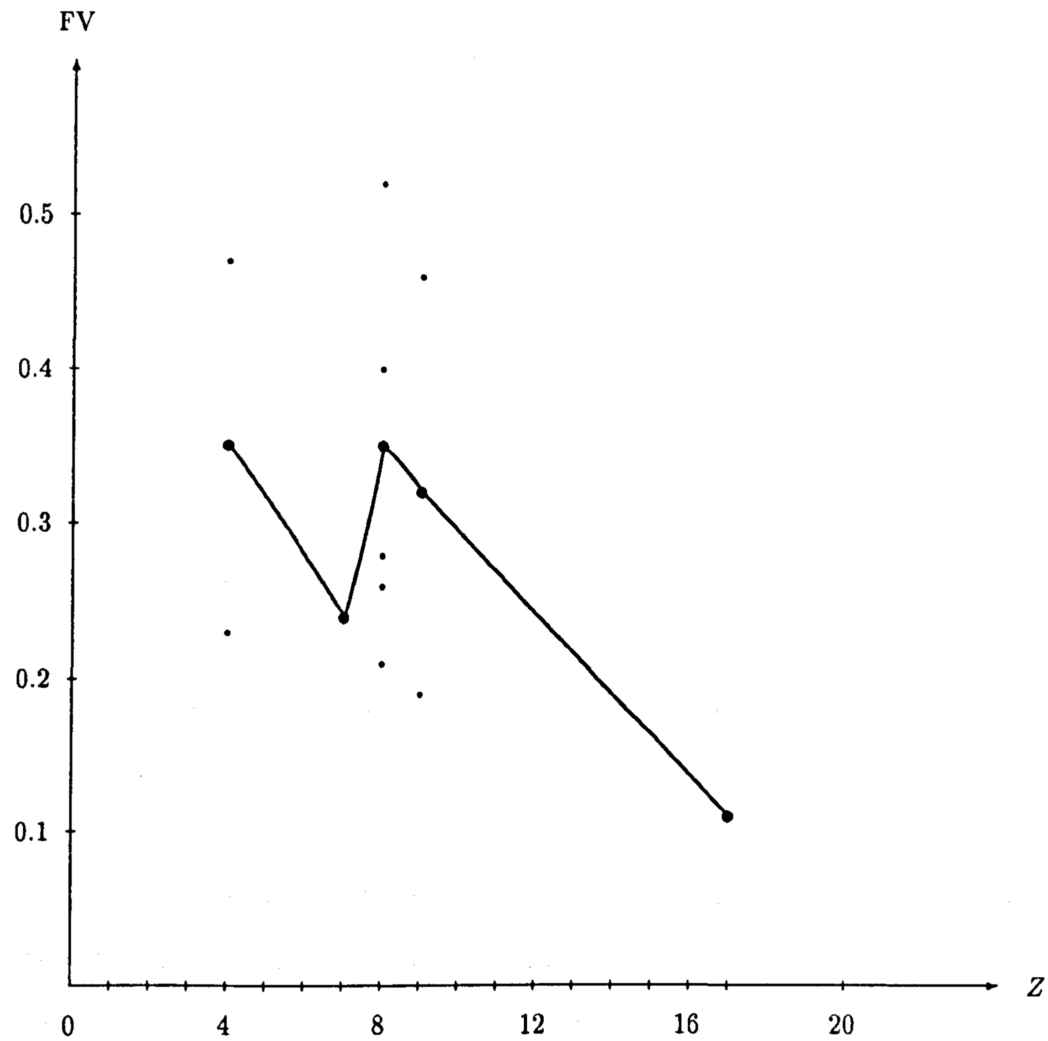

Figure 7

only one root;

(iii) its abstract form.

There was only one question with $Z = 7$, viz 1(iii) and its difficulty may due to the need to know the three forms of a complex number (i.e. algebraic, polar and trigonometric) and their relationships.

There was no significant difference in the mean FV of items with $Z < 6$ and the mean FV of those with $Z > 6$. This has already been explained.

<u>Using mean score to interpret data</u>

The mean score and the standard deviation has been calculated for each question. Table 10 shows the results, whereas Table 11 gives the average mean scores of items which have the same demand. These are illustrated in Figure 8.

Note that the pattern is quite different from that of facility value. The correlation between the mean score and the demand of question is positive, but not significantly. The more steps in the question, the more chance of collecting partial marks. There was no significant difference between the mean of mean scores of items with $Z < 6$ and the mean of mean scores of items with $Z > 6$.

In order to find out if there is any significant difference in the attainment in the two examinations, a comparison was made between the students' mean scores in the two examinations scaled to give a maximum possible mark of 100 (Table 12). A highly significant difference ($p < 0.0005$) between the two mean scores was found. The performance in the second exam was better than that in the first.

	1			2	3		4		5		6		
I	i	ii	iii		ii	iii	i	ii	a	b	i	ii	
												1	2
M	7.0	4.8	5.8	6.5	4.6	6.3	6.9	7.3	6.4	7.1	4.9	3.0	3.2
SD	3.8	4.5	3.7	2.5	4.0	4.3	3.8	2.9	4.4	3.9	4.3	4.3	4.3
Z	4	8	7	17	9	8	8	9	8	9	8	4	8

Table 10. The mean score and standard deviation for each item of the 2nd term exam of the "A" class.

Key: I, *M* and SD refer to item, mean score and standard deviation respectively.

Frq	2	1	6	3	1
mean	5.0	5.7	5.4	6.3	6.5
SD	4.1	3.7	4.3	3.6	2.5
Z	4	7	8	9	17

Table 11. The averages of mean scores of items which have the same demand.

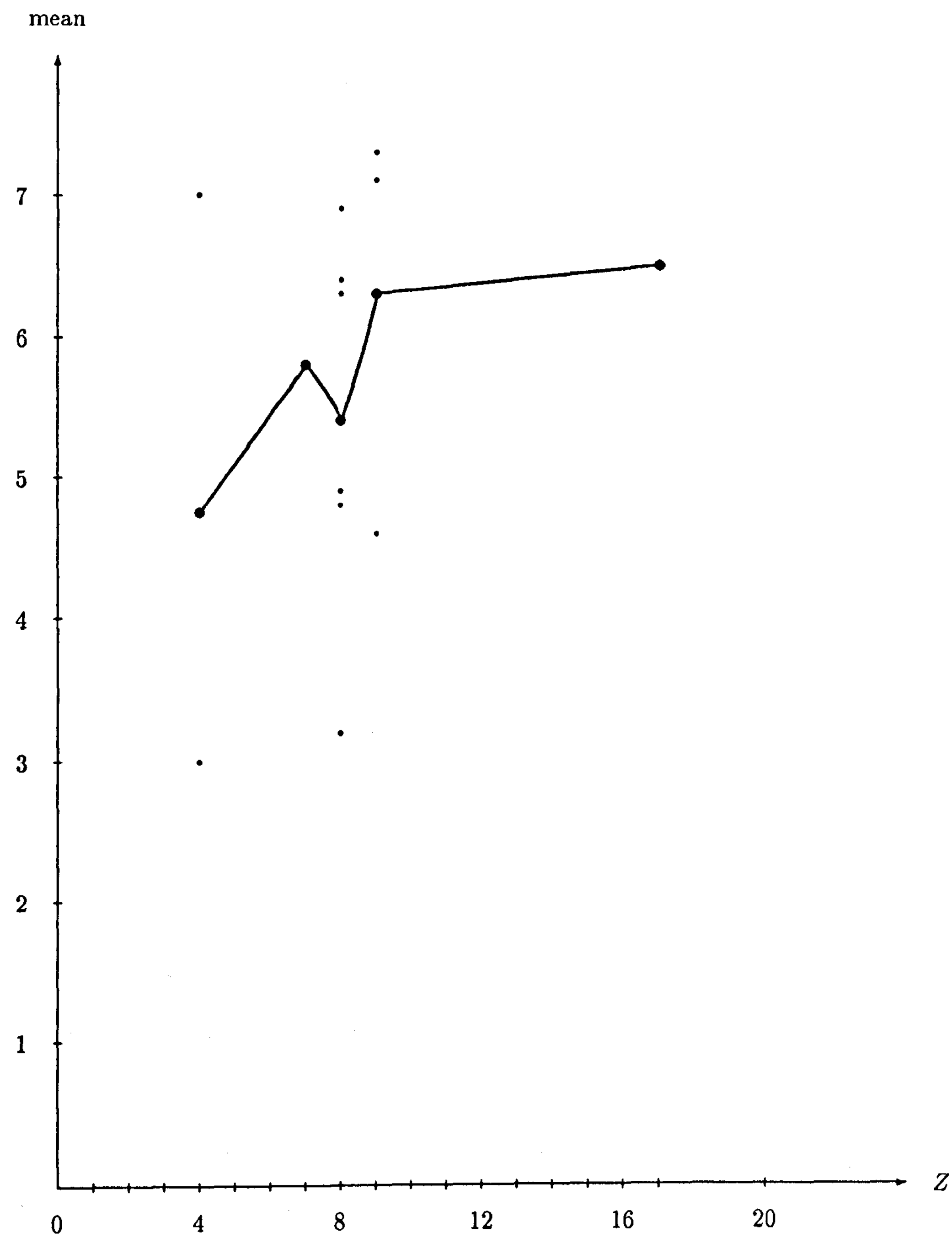

Figure 8

	First Exam	Second Exam
mean	48.6	59.2
SD	20.9	25.2

Table 12. Mean scores and standard deviations for the first and second examinations.

<u>The "B" class sample</u>

The sample contained 107 students. Following the same method as for the "A" class sample, the performance of students was traced during the two class examinations of the year 1987-88.

<u>The first term examination (November 1987)</u>

The exam was made up of twenty five items, but just twenty of them were used (one was theoretical and four of eight similar items were omitted for administrative reasons). The items were grouped into seven sets of complexity 2, 3, 4, 5, 6, 7 and 8. The analysis of the demand of the items has been done in the same way as with the "A" class. The complete November exam and analysis of some items can be found in Appendix 3.

The facility value for each item was calculated. This can be found in Table 13, while Table 14 shows the average facility values of items which have the same demand. These are illustrated in Figure 9.

As we expected, a negative correlation was found ($p < 0.01$) between facility value and the demand of items. So once again, "the

item		N	FV (%)	Z
1	i	61	57	3
	a	69	64	2
	b	39	36	2
	ii_a	77	72	3
	ii_b	67	63	4
	iii_a	89	83	5
	iii_b	82	77	3
	$iii_{b'}$	22	21	5
2	i	84	79	3
	ii	41	38	5
	iii_a	87	81	3
	iii_b	25	23	7
	iv	32	30	8
3	i	18	17	5
	ii	14	13	6
	iii	40	37	7
	iv	18	17	4
4	i_a	60	56	4
	i_c	60	56	5
	ii	23	21	6

Table 13. The FVs of items of the 1st term exam of the "B" class.

Frq	2	5	3	5	2	2	1
FV (%)	50	73	45	43	17	30	30
Z	2	3	4	5	6	7	8

Table 14. The average FVs of items which have the same demand.

subjects' performance decreases when the demand of questions increases". But note that, the performance of students on item 3(ii) with $Z = 6$, for example, was weaker than that on the item 2(iv) with $Z = 8$. The investigation of these items (whose thought steps have been analysed in Appendix 3) indicates that: the first item may require more effort than the second since the first tests both knowledge and understanding, while the second tests just knowledge in a straightforward manner.

The affect of the demand on the subjects' performance has been tested by comparing the mean FV of items with $Z < 6$ and that of items with $Z > 6$. The significant difference ($p < 0.01$) between two means supports the view that increasing demand of questions leads to a decrease in subjects' performance.

Using mean score to interpret data

The data has also been analysed in terms of mean score. Table 15 gives mean score and standard deviation for each item. Table 16 shows the average mean scores of items which have the same demand. These are plotted in Figure 10.

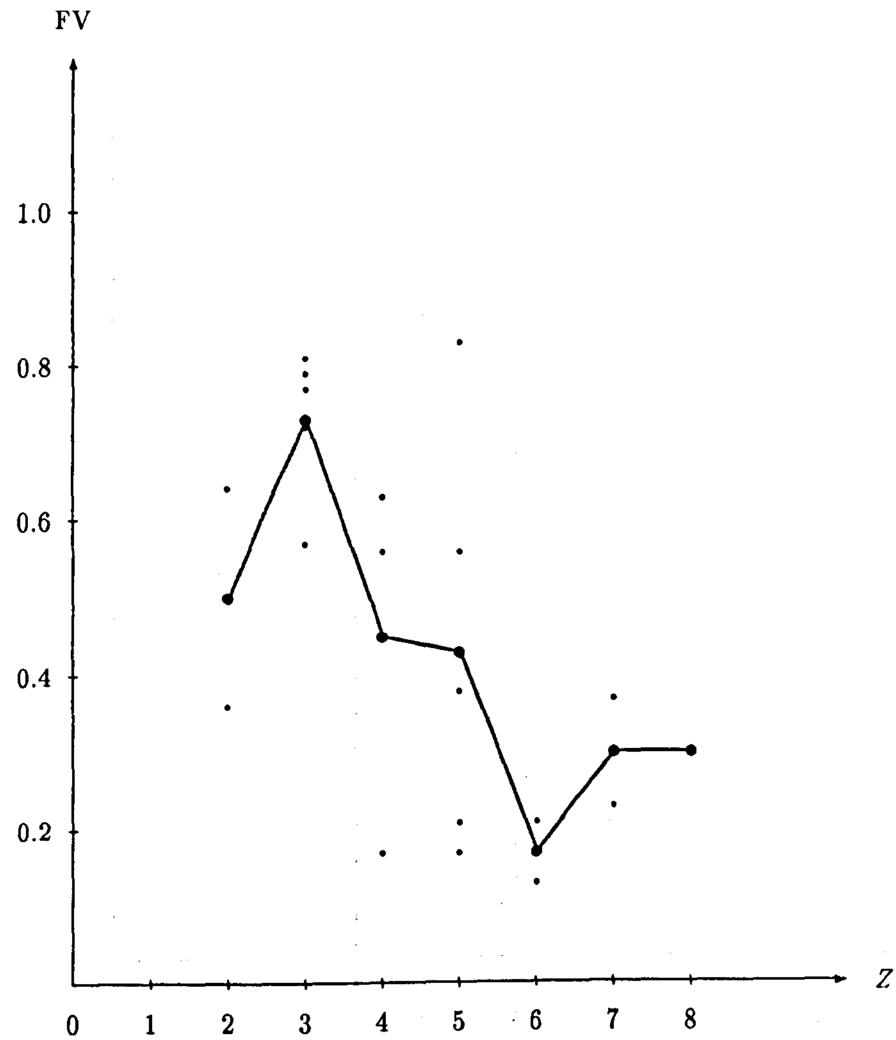

Figure 9

item		Z	mean	SD
1	i	3	7.3	3.7
	a	2	6.0	5.0
	b	2	4.0	5.0
	ii_a	3	8.5	3.0
	ii_b	4	7.3	3.7
	iii_a	5	9.0	3.0
	iii_b	3	8.0	4.0
	$iii_{b'}$	5	4.0	4.0
2	i	3	9.0	2.3
	ii	5	8.0	2.3
	iii_a	3	9.3	1.7
	iii_b	7	5.7	3.7
	iv	8	5.8	3.7
3	i	5	7.0	2.2
	ii	6	6.5	2.5
	iii	7	6.9	2.6
	iv	4	4.5	3.7
4	i_a	4	8.4	2.4
	i_c	5	8.0	2.8
	ii	6	5.0	3.6

Table 15. The mean scores and SDs (1st exam of the "B" class).

Frq	2	5	3	5	2	2	1
mean	5.0	8.4	6.8	7.2	5.8	6.3	5.8
SD	5.0	2.9	3.3	2.9	3.1	3.1	3.7
Z	2	3	4	5	6	7	8

Table 16. The averages of mean scores of items which have the same demand.

The pattern, in general, is similar to that for facility value, but the negative correlation obtained here is not significant. Again the question's demand has influenced students' performance: a significant difference ($p < 0.05$) was found between the mean of the mean scores of items with $Z < 6$ and that of mean scores of items with $Z > 6$.

<u>The second term examination (April 1988)</u>

Following the same procedure, the performance of the same sample of students in the April examination was tested, but some students were absent from the exam. The exam was made up of nineteen items, but one of them was theoretical and hence omitted from the investigation. The remaining eighteen items were grouped into seven sets of complexity 4, 5, 6, 7, 8, 9 and 14. The complete exam paper can be found in Appendix 4.

The facility value for each item was calculated. Table 17 gives the results, while Table 18 shows the average facility values of items which have the same demand. This is illustrated in Figure 11.

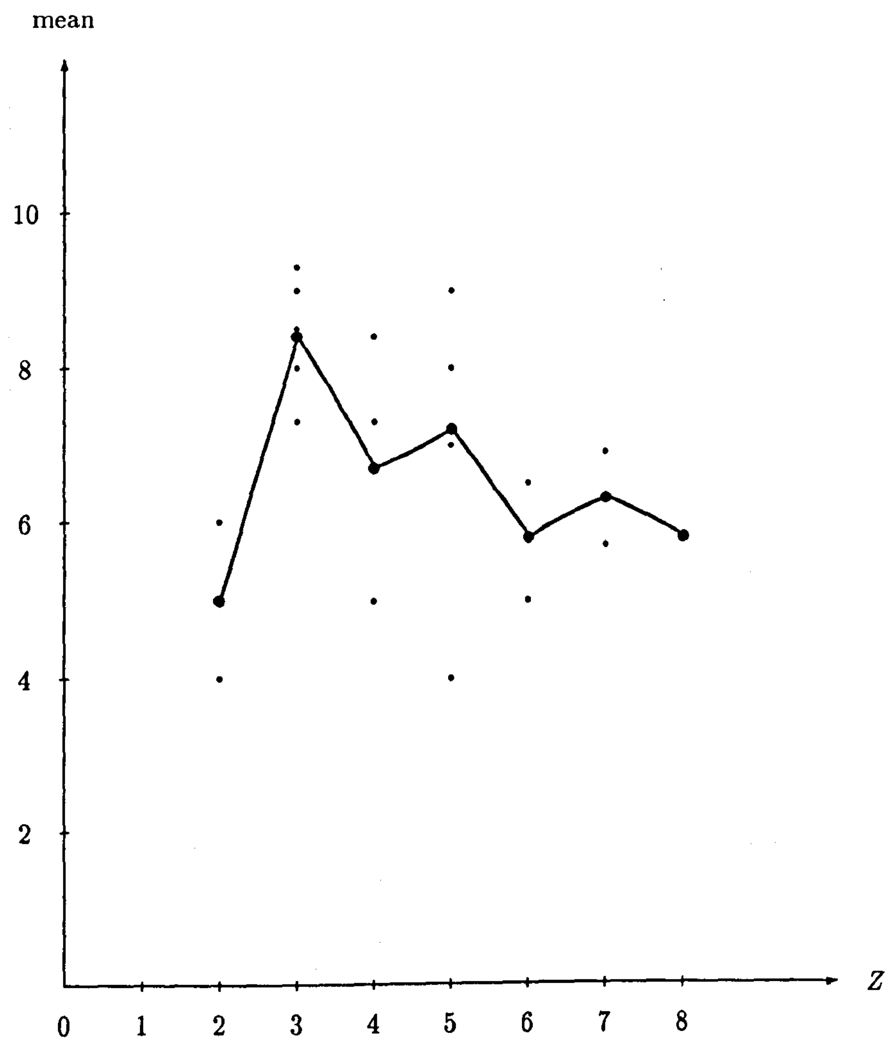

Figure 10

item	1		2			3		
	i	ii	i	ii	iii	i	ii	
							1	2
N	10	45	62	55	33	31	42	10
FV (%)	10	45	61	54	33	31	42	10
Z	14	5	6	4	7	7	5	8

Table 17. The FV for each item of the 2nd term exam of the "B" class.

item	4			5				6		
	i	ii	iii	ii	iii			i	ii	iii
					a	b	c			
N	21	28	34	55	38	35	33	30	40	27
FV (%)	21	28	34	54	38	35	33	30	40	27
Z	7	5	6	4	4	5	4	9	4	5

Table 17 (ctd).

Frq	5	5	2	3	1	1	1
FV (%)	44	36	48	28	10	30	10
Z	4	5	6	7	8	9	14

Table 18. The average FVs of items which have the same demand.

The interpretation of the results should be made in the light of the following observations.

(a) The test was composed of six questions, some involving a lot of information given in the form of text, and the last having abstract form.

(b) Some effort is required from the candidates to understand most of the items with $Z = 5$ before they solve them.

(c) The weakness of the candidates' performance on items with $Z = 7$ is entirely due to mathematical factors (e.g. difficulties with completing the square, finding partial fractions, and finding the general term of the series).

(d) The item with $Z = 8$ is not just hard mathematically, but also complex in its presentation.

(e) Both items with $Z = 6$ and $Z = 9$ seem to test mathematics concepts in a clear and short manner, and the method of solution was given.

(f) The item with $Z = 14$ is likely to be familiar to the students but it has many steps.

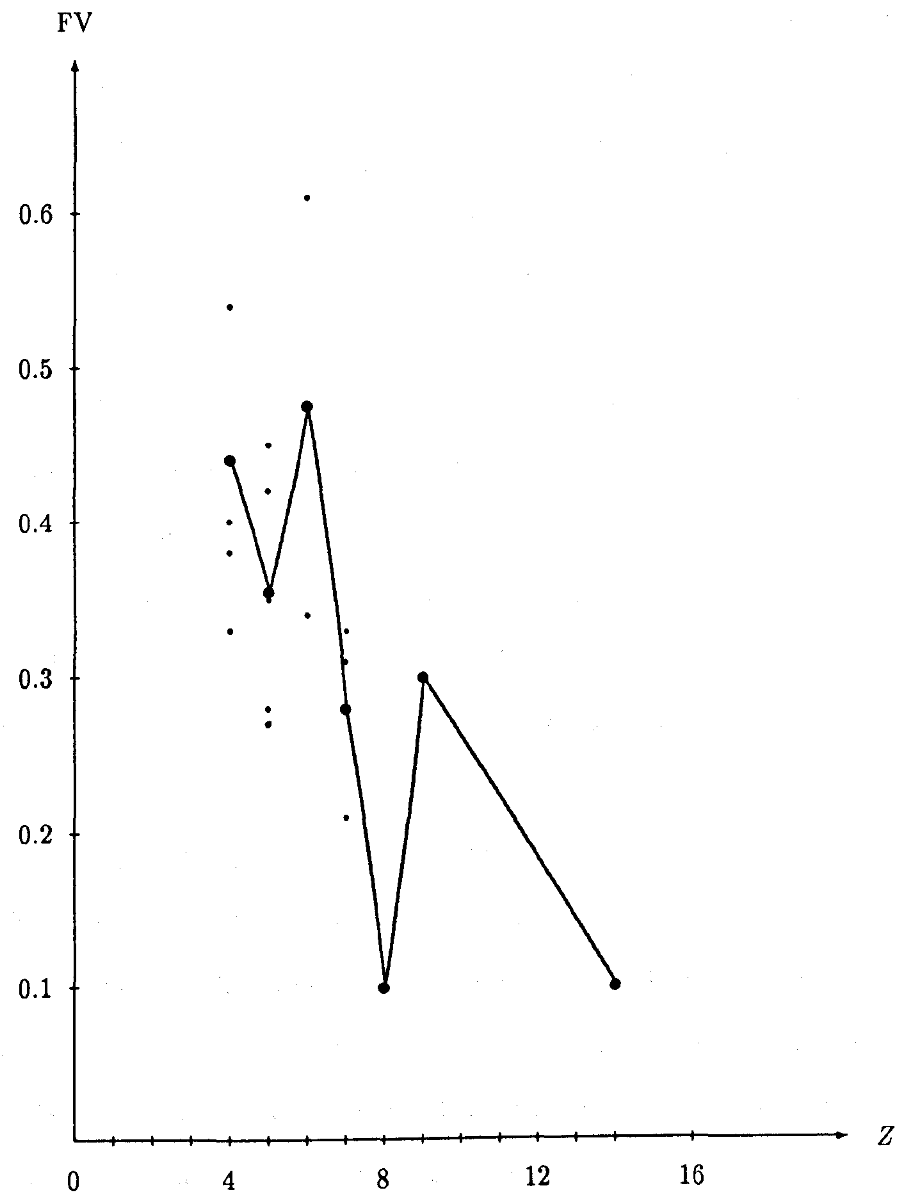

Figure 11

The subjects' good performance on items with $Z = 6$ and $Z = 9$, and relatively good performance on that with $Z = 14$ (considering its demand), can be explained by these observations. Other factors (such as the amount of information, the formulation, the testing of comprehension, etc.) may have influenced the performance on other items.

Note that a strong negative correlation between the facility value and the demand of items was obtained ($p < 0.01$), and a sharp drop of the curve comes after 6 on the Z-axis (Figure 11). Once again, the question's demand appears to affect students' performance in general: a significant difference ($p < 0.01$) between the mean score of FVs of items with $Z < 6$ and that of items with $Z > 6$ was obtained.

Using mean score to interpret data

The mean score has again been used to interprete the data. Table 19 gives the mean score and standard deviation for each item, while Table 20 shows the average mean scores of items which have the same demand. These are plotted in Figure 12. No negative correlation has been found between the two variables. This may due - as we mention earlier - to the nature of the mean score as a partially correct answer. Therefore, there is a chance of collecting partial marks from the individual steps.

No significant difference between the mean of the mean scores of items with $Z < 6$ and that of mean scores of items with $Z > 6$ was found.

In order to find out if there is any significant difference in the attainment in the two examinations, a comparison was made between the students' mean scores in the two examinations scaled to give a maximum

item	1		2			3		
	i	ii	i	ii	iii	i	ii	
							1	2
mean	6.2	5.4	7.5	6.0	5.0	7.0	5.3	1.9
SD	2.9	4.6	3.8	4.7	4.3	3.0	4.5	3.3
Z	14	5	6	4	7	7	5	8

Table 19. The mean score and standard deviation for each item of the 2nd term exam of the "B" class.

item	4			5				6		
	i	ii	iii	ii	iii			i	ii	iii
					a	b	c			
mean	5.5	4.8	5.7	6.3	5.0	5.0	3.3	7.4	5.8	3.8
SD	3.7	3.8	4.2	4.5	4.7	4.3	4.8	2.8	4.0	4.3
Z	7	5	6	4	4	5	4	9	4	5

Table 19 (ctd).

Frq	5	5	2	3	1	1	1
mean SD	5.3 4.5	4.8 4.3	6.6 4.0	5.8 3.6	1.9 3.3	7.4 2.8	6.2 2.9
Z	4	5	6	7	8	9	14

Table 20. The averages of mean scores of items which have the same demand.

possible mark of 100. Table 21 shows the results. There is a highly significant difference ($p < 0.0005$) between the two mean scores in favour of the first exam. This may be caused by the students' difficulties in the second examination as a result of the factors which we have already mentioned.

	First Exam	Second Exam
mean SD	67.9 16.1	55.6 22.7

Table 21. Mean score and standard deviation of the first and the second examinations.

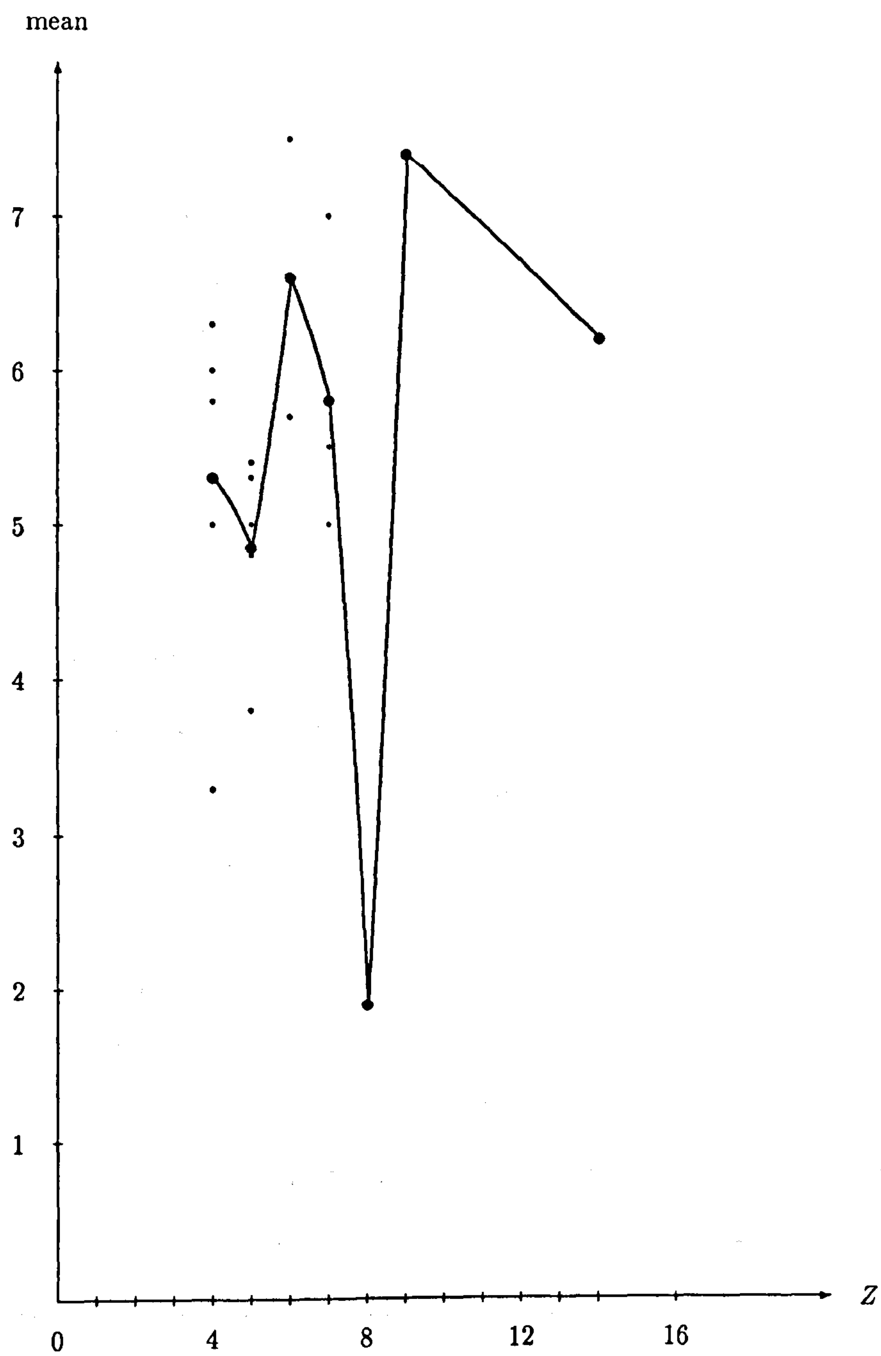

Figure 12

Application of the Holding-Thinking Space Model in Algerian School Mathematics

Following the method described for the research into tertiary level mathematics, in 1987, a sample of third-year mathematics pupils (aged 18) was selected from Amara Rachid Secondary School in Algeria. Pupils - at this level - need to sit three termly exams in their schools and the "Baccalaureat" exam. According to their results, they then enter a university, repeat their year at school or look for a job.

The pupils' progress in mathematics was traced for the three termly examinations in 1988-89, and their results in the "Baccalaureat" are discussed and analysed.

The first term examination (December 1988)

The sample was composed of 116 pupils who had to solve four problems which contained a total of fourteen items. The items were grouped according to their demand into six sets of complexity 2, 4, 5, 6, 8 and 9. In order to analyse the items into their demand, the procedure used with the Glasgow University exams was adopted here. The following two examples illustrate such an analysis.

Example 1

Find the set of remainders of (i) $(4)^n$, (ii) $(3)^n$ upon division by 7 for integers $n \geq 0$.

The two parts of this example have the same method of solution, therefore any separation of one from the other could lead perhaps to the same thought step being counted twice. The example can be analysed into the following steps:

step 1 identify the sequence of remainders upon division of $(4)^n$ by 7,

step 2 deduce the periodicity of $(4)^n$,

step 3 use the period to generate the result,

step 4 state the set of remainders upon division of $(4)^n$,

step 5 repeat the procedure for $(3)^n$,

step 6 state the set of remainders upon division of $(3)^n$.

Therefore, the *Z*-demand of this example is equal to 6 (thought steps).

Example 2

Let p be a prime number and a, b, c the integers which are represented by 7, 238, 1541 in base p respectively. Find p such that $c = cb$.

The thought steps of this example are:

step 1 write down 7, 238 and 1541 in base p,

step 2 substitute in $c = ab$,

step 3 deduce a value for p,

step 4 factorize the cubic equation into two factors,

step 5 show that the quadratic equation has no roots,

step 6 deduce that p has a unique value.

The demand of this item is equal to 6 (thought steps).

The complete December exam can be found in Appendix 5.

The facility value of each item of this exam was calculated and is tabulated in Table 22, while Table 23 gives the average facility values of items which have the same demand. These are illustrated in Figure 13.

A negative correlation between facility value and demand was obtained ($p < 0.05$): the subjects' performance decreases when the demand of the task increases. There is a significant difference ($p < 0.01$) between the mean score of FVs of items with $Z < 6$ and that of items with $Z > 6$.

Note that the pupils' performance on the item with $Z = 6$ was better than that on all other items except that with $Z = 2$. The explanation may be that, of the five items with $Z = 6$, viz 1(a), 1(b), 1(c), 3(a) and 4A(b), just the last one is difficult since it requires "proof by induction" which is usually hard for pupils. The remaining items need a familiarity with an appropriate algorithm and pupils usually apply these successfully. However, the pupils appear to have missed the algorithm for the question with $Z = 4$. Proof by induction occurs again in the question with $Z = 5$.

<u>Using mean score to interpret data</u>

The mean score has again been used to interpret data. Table 24 gives the mean score and standard deviation for each item of the December exam, while Table 25 gives the average mean scores of items which have the same demand. These are illustrated in Figure 14. It seems that the pattern is similar to that of facility value with the exception of the item with $Z = 8$, on which the performance was not poorer than that on items with $Z = 4$ and $Z = 5$. This may be due to the subjects' greater chance to collect partial marks from the

item	1 a	1 b	1 c	2	3 a	3 b	3 c
Frq	96	66	77	29	50	47	23
FV (%)	83	57	66	25	43	41	20
Z	6	6	6	8	6	2	8

Table 22. The FV for each item of the 1st term exam of the Algerian school.

item	4 Aa	4 Ab	4 Ac	4 Ba	4 Bb	4 Bc	4 Bd
Frq	99	36	34	33	39	15	7
FV (%)	85	31	29	28	34	13	6
Z	2	6	5	9	4	9	9

Table 22 (ctd).

Frq	2	1	1	5	2	3
FV (%)	63	34	29	56	23	16
Z	2	4	5	6	8	9

Table 23. The average FVs of items which have the same demand.

item	1 a	1 b	1 c	2	3 a	3 b	3 c
mean	9.3	7.0	7.5	5.7	5.6	4.8	2.6
SD	1.8	4.0	3.8	3.8	4.4	4.8	4.1
Z	6	6	6	8	6	2	8

Table 24. Mean score and SD for each item of the 1st term exam of the Algerian school. (The table continues on the next two pages.)

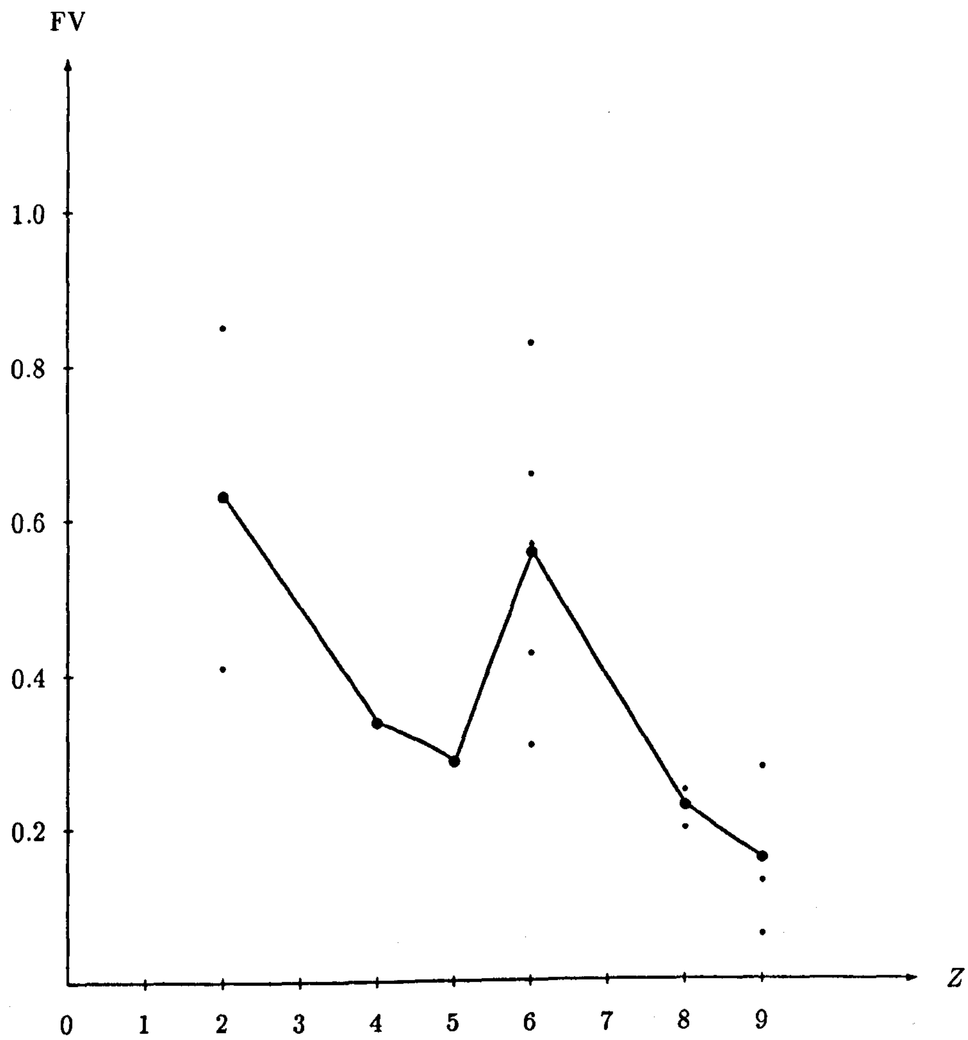

Figure 13

item	4						
	Aa	Ab	Ac	Ba	Bb	Bc	Bd
mean	9.0	4.0	3.0	4.8	4.0	2.9	1.0
SD	3.0	4.5	4.5	4.3	4.5	3.9	2.5
Z	2	6	5	9	4	9	9

Table 24 (ctd).

Frq	2	1	1	5	2	3
mean	6.9	4.0	3.0	6.7	4.2	2.9
SD	3.9	4.5	4.5	3.7	3.9	3.6
Z	2	4	5	6	8	9

Table 25. The average mean scores of items which have the same demand.

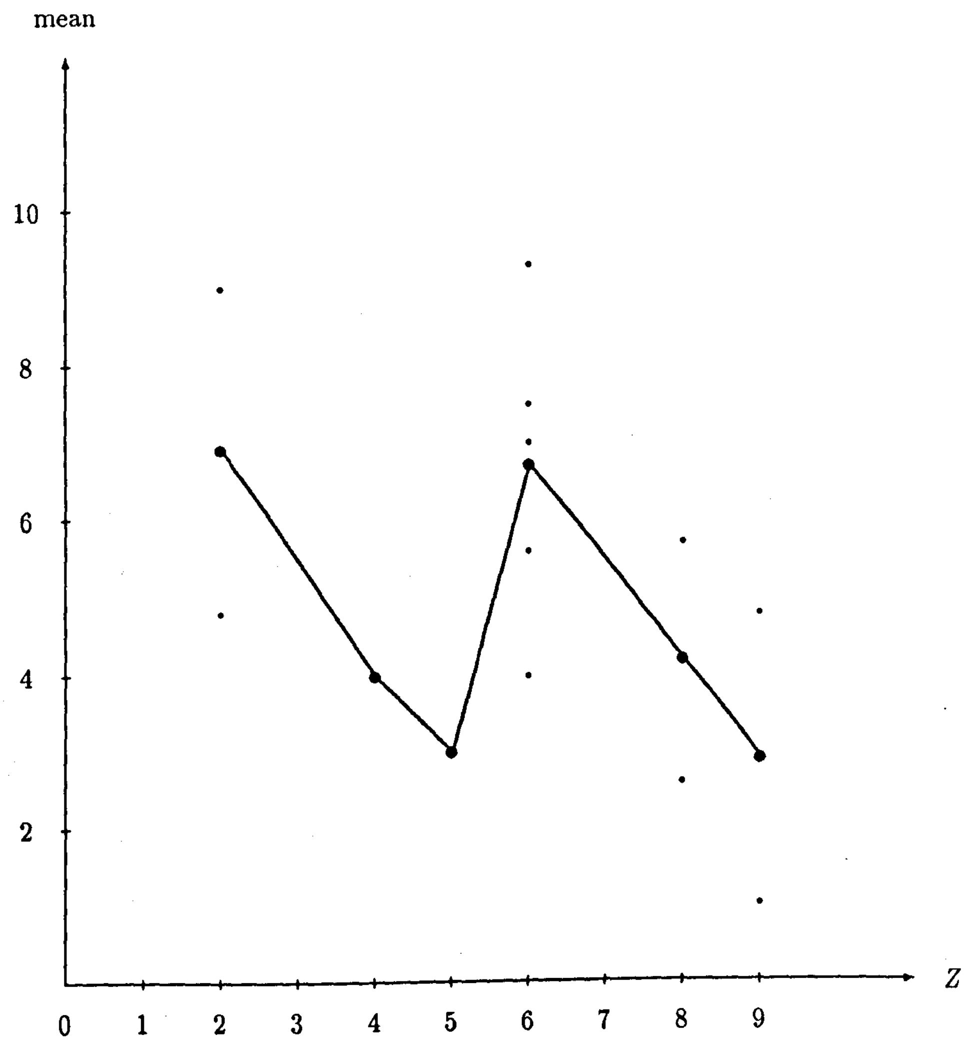

Figure 14

individual steps of the item with $Z = 8$ than those with $Z = 4$ or $Z = 5$. Note also that a negative correlation between mean score and the demand was obtained (although it was not significant) and, in general, the demand affects the subjects' performance: a significant difference ($p < 0.1$) between the mean of mean scores of items with $Z < 6$ and that of mean scores of items with $Z > 6$ was obtained.

The second term examination (March 1989)

The sample size was reduced to 106 since ten pupils were absent from the exam. The subjects had to solve two questions and one problem (or sixteen items) in the exam. These items were analysed into their demand and grouped into eight sets of complexity 2, 5, 6, 7, 8, 9, 10 and 11. The complete March exam can be found in Appendix 6.

The facility value was calculated for each item. Table 26 gives the results, while Table 27 shows the average facility values of items which have the same demand. These are illustrated in Figure 15.

Before any interpretation of the results, I think the following points should be noted.

Firstly, the main difficulty with the first question of the exam is that it contains five parts, each of them (except the first) depending on the previous parts. Therefore, the load may increase when we move from one part to the next one (this can be seen clearly in Table 26). We find the subjects' performance on the item with $Z = 6$ is poorer than that on the item with $Z = 11$, which occurs much earlier in the question. Moreover, the third part has been estimated to have two thought steps (determine the coordinates of points which represent the solution, and plot them). To solve this part, you need

item	1					2			
	Ia	Ib	Ic	IIa	IIb	1	2	3	4
Frq	92	59	53	2	6	78	52	24	8
FV (%)	87	56	50	2	6	74	49	23	8
Z	7	11	2	10	6	2	7	9	9

Table 26. The FV for each item of the 2nd term exam of the Algerian school.

item	3						
	1	2a	2b	3	4a	4b	4c
Frq	72	58	3	35	65	33	8
FV (%)	68	55	3	33	60	31	8
Z	9	9	6	5	8	5	9

Table 26 (ctd).

Frq	2	2	2	2	1	5	1	1
FV (%)	62	32	5	68	60	32	2	56
Z	2	5	6	7	8	9	10	11

Table 27. The average FVs of items which have the same demand.

the results of the first two parts. Therefore, it is not surprising to find little difference between the performance on items of such different demand as $Z = 11$ and $Z = 2$. A question which may arise here is: why do we not deal with this question as a whole rather than as parts? The simple answer is that if we do so the demand of this question reaches 36 (thought steps) and I believe that a question with dependent parts in which the solution of each part needs the result of the previous one may need further analysis of its demand.

Secondly, even though parts of the second question are concerned with the concept of barycentre, pupils had the opportunity to deal with a part without the necessary of solving the previous one.

Thirdly, the problem has seven parts in which the first was familiar to the students, while the second required the definitions of both continuity and differentiability. The load of the third part (with $Z = 6$) may be due to problem of understanding modulus, whereas the difficulty in part four (with $Z = 5$) was clearly the subject

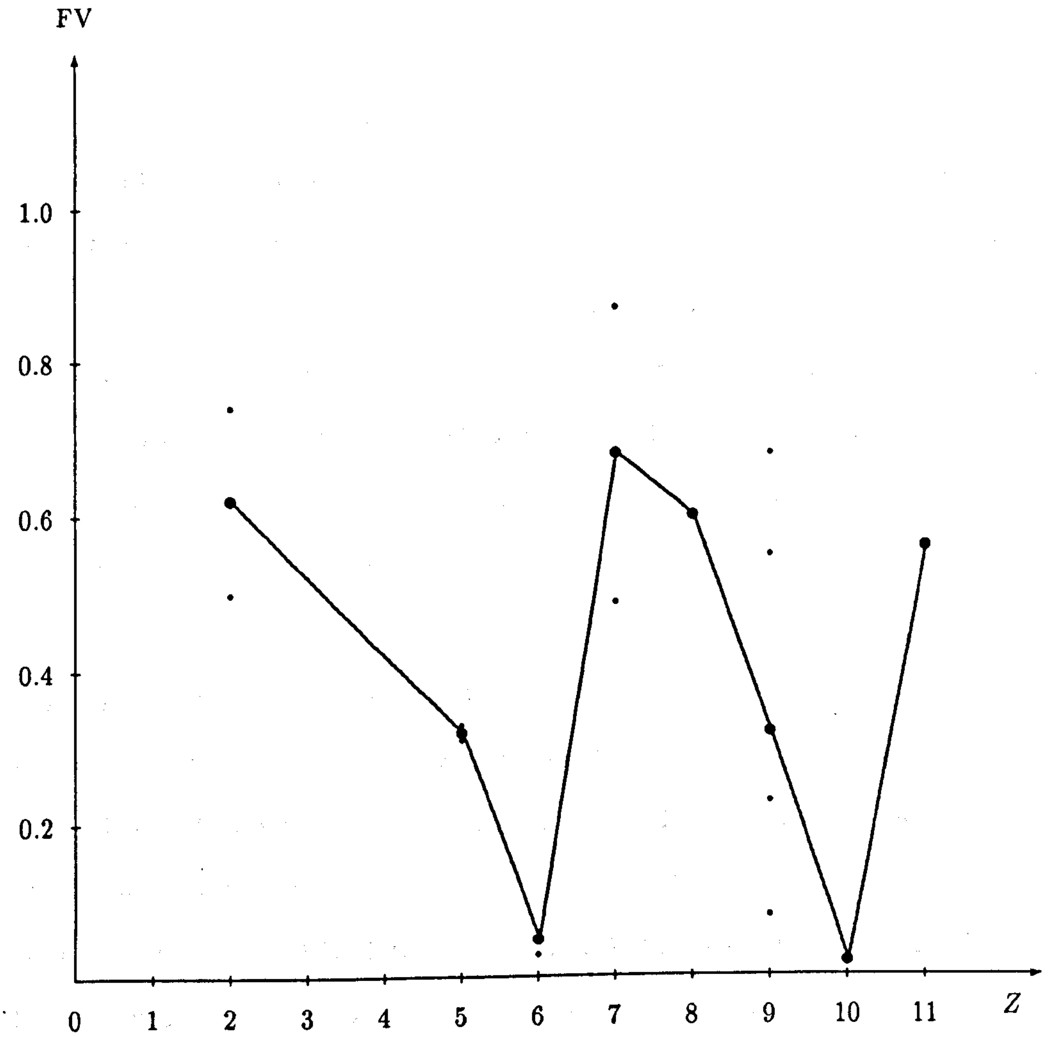

Figure 15

matter (bijective and inverse functions). It seems to me that because pupils were familiar with the evaluation of integrals, the high performance on part five (with $Z = 8$) was expected. While the difficulty of part six (with $Z = 5$) may arise from the limits which, in general, are hard to handle.

A negative correlation between facility value and demand was obtained but it was not significant since, as remarked above, subjects succeeded in some items which have a relatively high demand, while they had little success on others which were characterized by a low demand. This can been seen clearly in Figure 13. No significant difference was found between the mean FV of items with $Z < 6$ and the mean FV of those with $Z > 6$.

Using mean score to interpret data

In order to interpret data by another factor, a mean score and standard deviation was calculated for each item. Table 28 gives the results, whereas the average mean scores of items which have the same demand are given in Table 29. These are illustrated in Figure 16.

In general, the pattern obtained with the facility value factor was repeated here in terms of a negative correlation; some items were characterised by a high (low) achievement and also had a high (low) demand, and no significant difference was found between the mean of mean scores of items with $Z < 6$ and that of mean scores of items with $Z > 6$.

item	1					2			
	Ia	Ib	Ic	IIa	IIb	1	2	3	4
mean	8.9	6.7	5.2	2.2	2.6	7.6	5.8	3.9	2
SD	2.8	4.1	5.2	2.9	2.8	4.2	4.5	4.1	3.3
Z	7	11	2	10	6	2	7	9	9

Table 28. Mean score and standard deviation for each item of the 2nd term exam of the Algerian school.

item	3						
	1	2a	2b	3	4a	4b	4c
mean	8.9	6.9	3.8	4.8	6.4	4.4	1.8
SD	1.9	3.9	2.9	4.2	4.7	4.4	3.2
Z	9	9	6	5	8	5	9

Table 28 (ctd).

Frq	2	2	2	2	1	5	1	1
mean SD	6.4 4.7	4.6 4.3	2.7 2.8	7.3 3.6	6.4 4.7	4.7 3.3	2.2 2.9	6.7 4.1
Z	2	5	6	7	8	9	10	11

Table 29. The average mean scores of items which have the same demand.

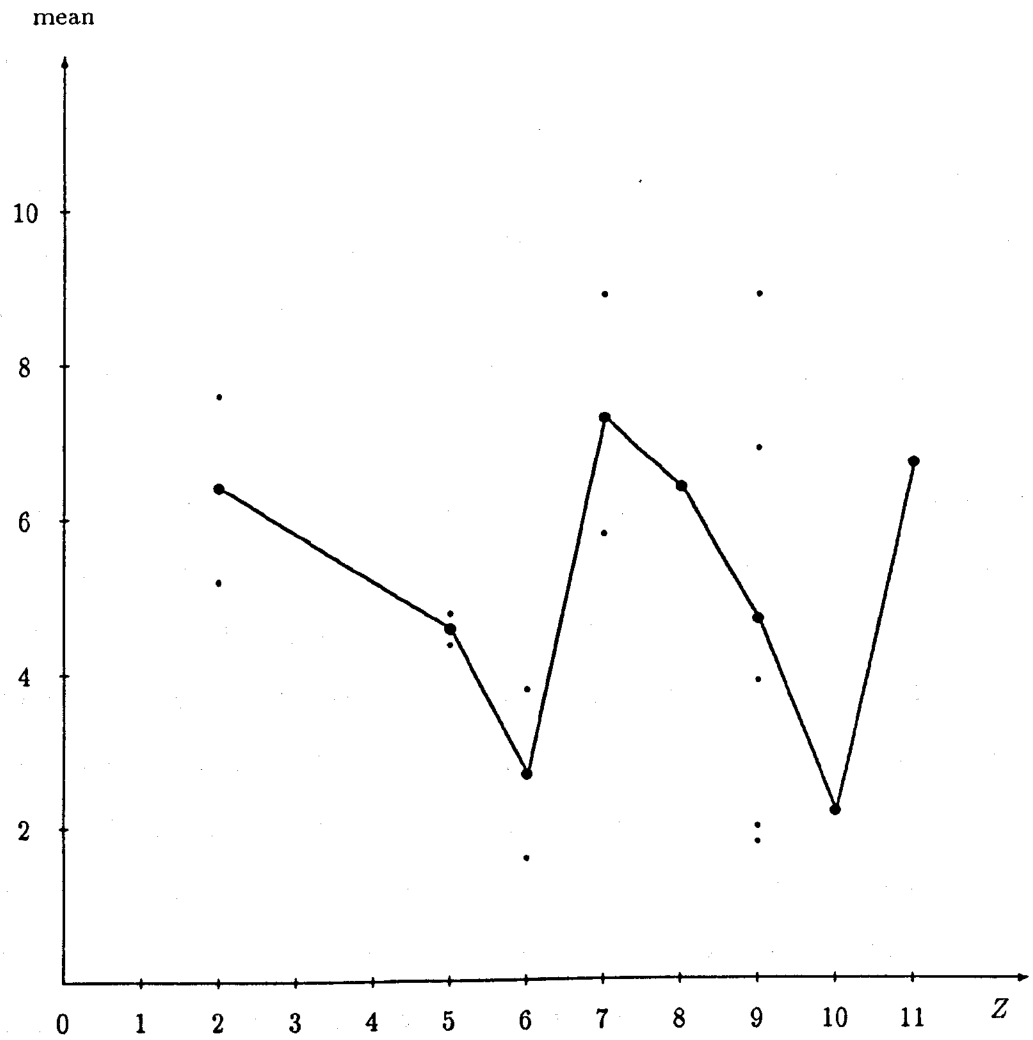

Figure 16

The third term examination (May 1989)

Because one class did not keep the scripts in the school, I was obliged to exclude this class from the study even though I have their total marks but no detailed marks for each item. Consequently, the sample size reduces to 66. The exam was composed of two questions (i.e. 1 and 2) and one problem with five parts (i.e. I, II, III, IV and V), but the number of items involved in this exam was thirty. These items were analysed into their demand and grouped into ten sets of complexity 2, 3, 4, 5, 6, 7, 8, 9, 10 and 13. The complete May exam can be found in Appendix 7.

Following the method which I applied to the first two examinations, the facility value of each item was calculated. Table 30 shows the results, whereas Table 31 gives the average facility values of items which have the same demand. These are illustrated in Figure 17. No significant difference was found between the mean FV of items with $Z < 6$ and the mean FV of those with $Z > 6$. We expect the low performance in the second of these groups since it is characterised by the high demand of each of its elements. But the low performance on some items in the first group (and the items with $Z = 6$) needs explanation.

Item with $Z = 4$ (i.e. II.2)

The problem about determining the demand of dependent parts has again emerged in this item (see also the discussion of the second exam). The demand of this item was identified by counting thought steps used after the solution of the previous part (i.e. II.1). But in the situation in which pupils start from the second part rather than the first) the demand of this item is much higher than 4.

item		frq	FV (%)	Z
1	1	63	95	2
	2(a)	44	67	3
	2(b)	44	67	9
	2(c)	15	23	10
	3	8	12	7
2	1	23	35	10
	2_a	10	15	5
	2_b	0	0	13
I	1	37	56	7
	2	6	9	5
	3(a)	16	24	5
	3(b)	10	15	2
	4	24	36	3
II	1	31	47	10
	2	20	30	4
	3	1	2	6
	4	11	17	2

Table 30. The FV for each item of the 3rd term exam of the Algerian school. (The table continues on the next page.)

item		frq	FV (%)	Z
III	1	53	80	2
	2(a)	19	29	6
	2(b)	47	71	7
	2(c)	33	50	6
	2(d)	1	2	9
IV	1	14	21	8
	2	27	41	8
	3	0	0	10
	4(a)	10	15	5
	4(b)	18	27	7
V	1	2	3	9
	2	4	6	10
	3	4	6	8

Table 30 (ctd).

Frq	4	2	1	4	3	4	3	3	5	1
FV (%)	52	52	30	16	27	42	23	24	23	0
Z	2	3	4	5	6	7	8	9	10	13

Table 31. The average FVs of items which have the same demand.

Items with $Z = 5$

There are four items with $Z = 5$, viz 2.2_a, I.2, I.3(a) and IV.4(a).

The first item involves a proof by induction which is not easy to master.

The difficulty of the second item may due to the lack of precision in the question about the transformations.

The third item was in general form and pupils found it difficult to apply the condition: "tangent is parallel to the x-axis".

The fourth item was again imprecise in its description of the three transformations.

Items with $Z = 6$

There are three items with $Z = 6$, viz II.3, III.2(a) and III.2(c), but the last item is acceptable.

The main difficulty of the first item is that to find the expression for b_n in terms of n requires insight to see the

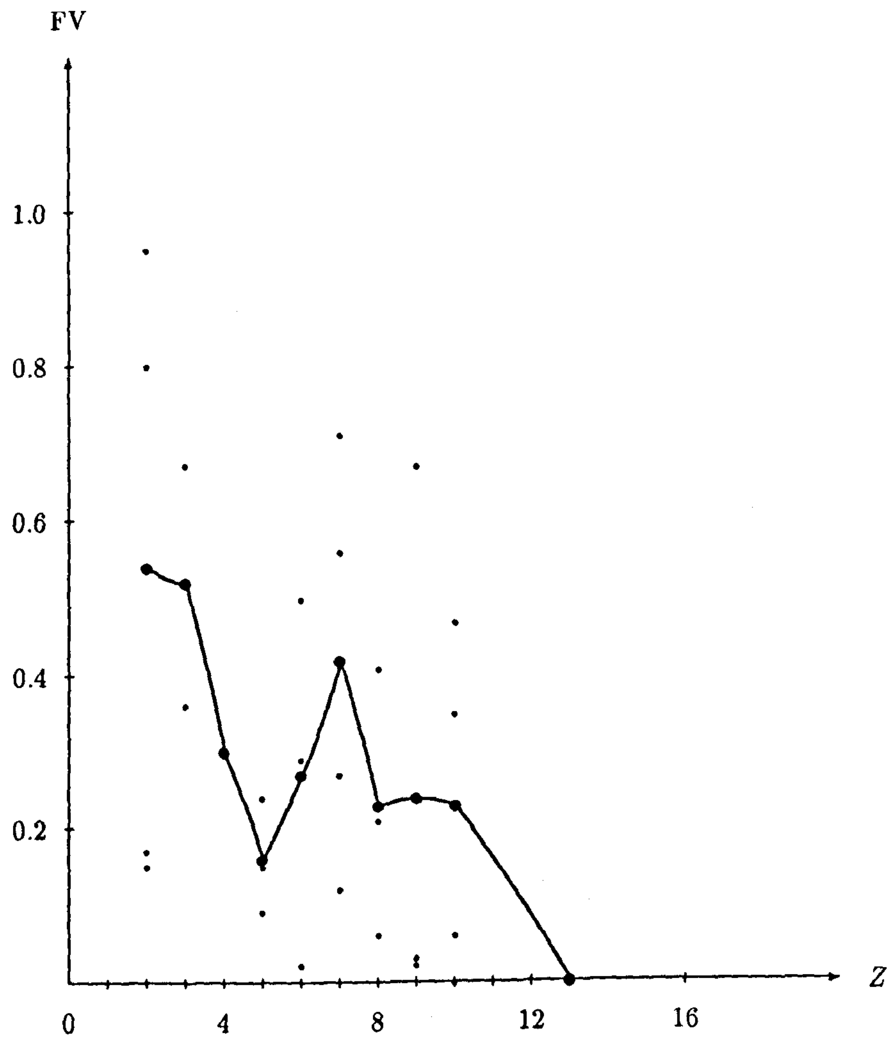

Figure 17

appropriate strategy.

The difficulty of the second may be in recalling the definition of invariant points or formulating the final result in a way which shows there is nothing missing from the solution (particularly the origin).

Note that the performance on the items with $Z = 7$ was in general better than that on the other items. However, there were four items with this demand, viz 3, I.1, III.2(b) and IV.4(b), and the middle two clearly have high facility values since in the first case the item was familiar and the other involved a routine algorithm.

A significant correlation between facility value and demand was obtained ($p < 0.05$) but the demand of items was, in general, affected by the above factors.

Using mean score to interpret data

Table 32 gives the mean score and the standard deviation for each item of this exam, while Table 33 shows the average mean scores of items which have the same demand. These are plotted in Figure 18.

It seems that the pattern is likely to be similar to that of facility value. A negative correlation was obtained (nearly significant at the 5% level, but significant at the 1% level in case of the average mean scores) and no significant difference was found between the mean of mean scores of items with $Z < 6$ and that of mean scores of items with $Z > 6$.

item		mean	SD	Z
1	1	9.6	2.0	2
	2(a)	6.6	4.8	3
	2(b)	7.5	3.8	9
	2(c)	3.7	4.2	10
	3	2.2	3.2	7
2	1	4.9	4.1	10
	2_a	2.5	3.7	5
	2_b	0.2	0.8	13
I	1	8.6	2.2	7
	2	1.7	3.2	5
	3(a)	2.4	4.4	5
	3(b)	1.5	3.6	2
	4	4.4	4.6	3
II	1	4.0	4.0	10
	2	4.0	4.4	4
	3	1.5	2.5	6
	4	1.7	3.7	2

Table 32. Mean score and standard deviation for each item of the 3rd term exam of the Algerian school. (The table continues on the next page.)

item		mean	SD	Z
III	1	8.0	3.9	2
	2(a)	4.2	4.2	6
	2(b)	7.2	4.4	7
	2(c)	5.2	5.2	6
	2(d)	1.2	2.4	9
IV	1	4.6	3.7	8
	2	5.4	4.2	8
	3	1.9	2.5	10
	4(a)	2.6	3.8	5
	4(b)	2.7	4.4	7
V	1	0.6	1.9	9
	2	1.3	2.8	10
	3	0.8	2.6	8

Table 32 (ctd).

Frq	2	2	1	4	3	4	3	3	5	1
mean	5.2	5.5	4.0	2.3	3.6	5.2	3.6	3.1	3.2	0.2
SD	3.3	4.7	4.4	3.8	3.9	3.5	3.5	2.7	3.5	0.8
Z	2	3	4	5	6	7	8	9	10	13

Table 33. The average mean scores of items which have the same demand.

In order to find out if there is any difference in the attainment in the three examinations, a comparison was made between pupils' mean scores in the three examinations scaled to give a maximum possible mark of 100. The pupils' results in the Baccalaureat examination were also analysed.

Table 34 gives mean score and standard deviation for the subjects' scores in the three examinations. Note that there was no significant difference between the first and third exam, while such a difference was found between the second and each of the other two ($p < 0.0005$ in each case). The explanation may be that the second term is usually longer than the first or the third terms. Therefore, pupils may have more time to revise and organise their second term work. The difficulty of the third term examination may be due to its design as a preliminary exam to the Baccalaureat exam.

Table 35 shows the results of the pupils in the Baccalaureat

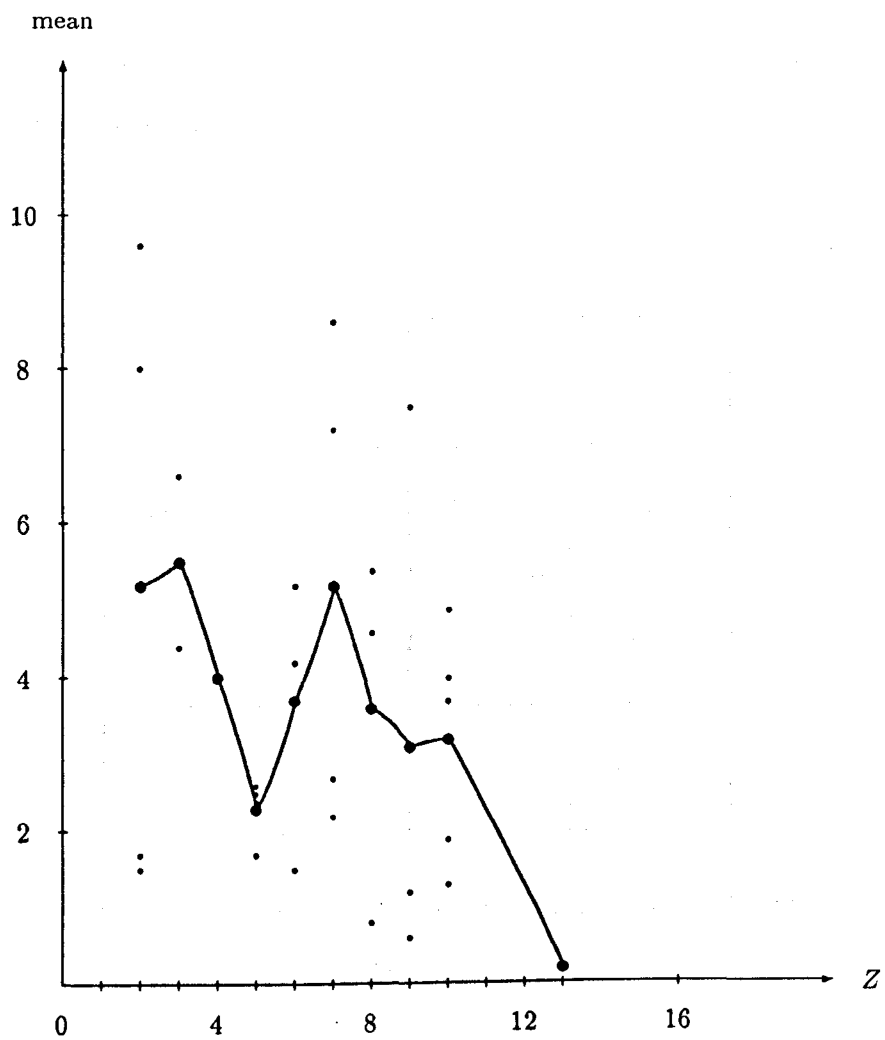

Figure 18

examination in session 1988-89. Even though the success in this certificate was around 59% which is very encouraging compared with previous years, there was no significant difference between the number of pupils who pass and the number who fail. This shows the difficulty of passing such exam.

	mean	SD
1st term exam	39.1	17.7
2nd term exam	52.4	21.9
3rd term exam	38.7	15.6

Table 34. Mean scores and standard deviations of the three termly examinations.

pass	fail
59%	41%

Table 35. The results of the Baccalaureat examination in session 1988-89 in the Algerian school.

CHAPTER FOUR

The Influence of Field Dependent-Independent Cognitive Style on Mathematics Performance

Field dependent-independent cognitive style could be interpreted in the basis of the differentiation theory which refers to the complexity of the structure of the psychological system in which the organism has to function [111]. During the presentation of Neo-Piagetian theory, we reported that Pascual-Leone [27] characterized field dependent subjects by their habitually low mental processes and highly influenced by the perceptual field rather than the task instructions. Whereas, field independent subjects are assumed to be the opposite. Witkin [93] distinguished the two categories in terms of their ability to overcome an embedding context which appears distinct from their ability to overcome the effect of distracting fields and also in terms of their relying an external or internal referents in processing information [111]. Moreover, it was considered [112] that field independent subjects are people who have the ability to break up an organized perceptual field and separate an item from its context. While field dependent subjects were taken to be the opposite.

Note that field independent people are more efficient than those who are field dependent in their ability to select relevant information. Therefore, the performance of the former is likely to be better and more accurate than the latter in both higher order and lower order tasks [113]. This may explained by the quality of recalled information since the poor performance of field dependent pupils may due to their difficulty of remembering the appropriate information. This opinion has been stressed by Riding and Person [114] who considered the way in which information is both analysed and represented in memory is the reason for superiority - in art performance - of field independent subjects over field dependent ones. But Goodenough [115] noted that each category can perform better than the other under certain conditions. Therefore, the difference between the two groups may be due to the process they employ rather than its effectiveness.

We have already mentioned that the size of subjects' capacity has been adopted as 6 ± 2. Therefore, people can be divided into three categories according to their capacity size X: low capacity (i.e. $X = 4$ or $X = 5$), average capacity (i.e. $X = 6$) and high capacity (i.e. $X = 7$ or $X = 8$). Generally low capacity people are field dependent, while high capacity people are field independent, but they are by no means perfectly correlated. It was reported [2] that in each X - space group, it seems that field independent subjects have the ability to obtain higher scores in the same examinations than the field dependents. Moreover, the mean scores of the $X = 7$ subjects in each examinations are higher than the $X = 6$ subjects, while both are higher than the $X = 5$ subjects. These findings support El-Banna's work [5]. However, according to Johnstone [116], there is a strong relationship between the subject score and the ability to ignore irrelevant distracting material. The field dependent person has difficulty in separating relevant material from irrelevant. If he takes in irrelevant material (along with the relevant) he does not have the capacity to cope with the problem. As a result of this, lowest performance will occur for low capacity field dependent subjects. On the other hand, people who are of high capacity but field dependent can tolerate some irrelevant material and still solve the problem. The most likely to be successful would be the high capacity field independent people because they have plenty of space and handle only or mainly relevant ideas.

According to Sharrat [1], field independent students tend to be better in mathematics and science, also they are more successful in imposing structure on an unstructured setting. Field dependent students, however, are more adept at interpersonal skills although they seem to be handicapped by an unstructured learning situation. The difference in ability between the two groups appears to be related to at least some aspects of discovery learning in mathematics. It was found that [117] a significant disordinate interaction between field independence and the level of guidance of instruction in mathematics. Field independent students did significantly better when the treatment provided minimal guidance, whereas the field dependent students seemed to learn more under conditions of maximal guidance.

In order to find out the influence of field dependent-independent cognitive style on the performance of the students in mathematics, the researcher applied a **Hidden Figures Test** (HFT) on the "A" class sample of first year university students and on the Algerian school sample (after the translation of the instructions of this test). In the case of "B" class, the test was already used by other researchers. The description of the test can be found in Appendix 8.

The "A" class sample

The first term examination (November 1987)

(a) The sample

The number of students of "A" class who attempted the (HFT) was 107. Their distribution of the total scores has been illustrated in Figure 19. This distribution divides the sample into three categories according to subjects' marks in the (HFT): subjects were considered as field dependent or field independent or field intermediate if their marks are below one standard deviation or above one standard deviation or between them respectively ([92], [96]). As a result of this classification, Table 36 gives the three groups' size.

(b) The results

The comparison between the students attainment in the November exam and their degree of field dependence-independence has been made. Table 37 shows the mean score and standard deviation of three groups. From the results of Table 38, the only significant difference which has been obtained is between field dependent-intermediate groups. A low correlation between scores and field dependent-independent marks has been found.

The second term examination (April 1988)

(a) The sample

The sample size has been reduced to 101. The scores of the students has been illustrated in Figure 20, and the three groups have been identified in the same manner as above. The size of the groups is shown in Table 39.

(b) The results

The mean score and the standard deviation has been calculated for each group (Table 40). No significant difference in means has been found (Table 41). A low negative correlation has been found between

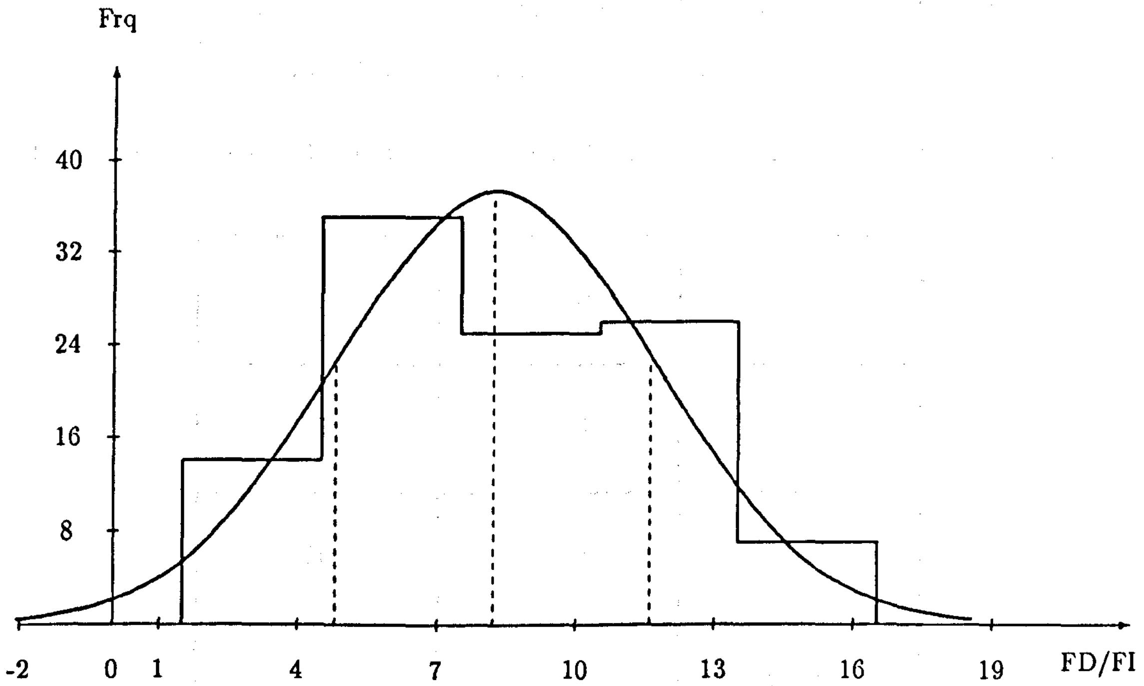

Figure 19

range 13

median 8

mean 8.2

SD 3.4

	FD	FInt	FI
size	19 (18%)	65 (61%)	23 (21%)

Table 36. Classification into FD/FI groups (1st exam "A" class).

group	mean	SD
FD	41.3	17.5
FInt	50.7	21.6
FI	45.6	23.5

Table 37. mean and SD of FD/FI groups (1st exam "A" class).

	FD	FInt	FI
FD	1		
FInt	S (5%)	1	
FI	NS	NS	1

Table 38. The significant difference in means (1st term exam/A class)

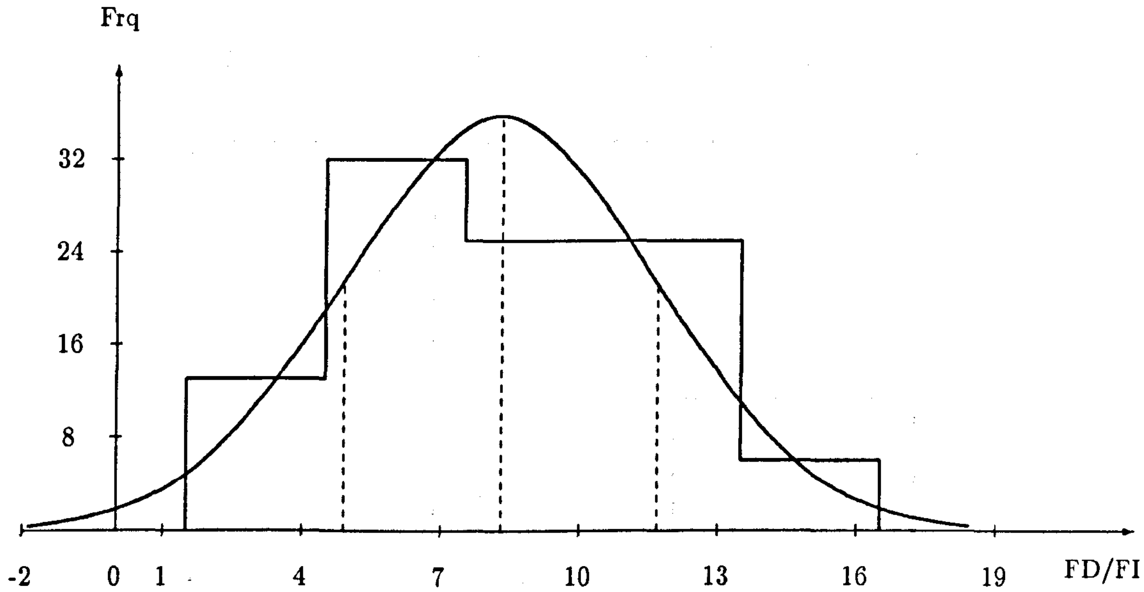

Figure 20

range 13

median 8

mean 8.3

SD 3.4

	FD	FInt	FI
size	17 (17%)	62 (61%)	22 (22%)

Table 39. Classification into FD/FI groups (2nd exam "A" class).

group	mean	SD
FD	53.7	26.8
FInt	62.3	23.1
FI	53.1	28.5

Table 40. Mean and SD of FD/FI groups (2nd exam "A" class).

	FD	FInt	FI
FD	1		
FInt	NS	1	
FI	NS	NS	1

Table 41. The significant difference in means (2nd exam "A" class).

the sample marks and the field dependent-independent marks.

The "B" class sample

The first term examination (November 1987)

(a) The sample

The sample size of "B" class who attempted the (HFT) was 85. The distribution of the scores has been illustrated in Figure 21. The three groups have been identified and their size is given in Table 42.

(b) The results

Table 43 gives mean score and standard deviation for each group. Note that no significant difference in means has been found (Table 44). A low correlation between scores of both the exam and the (HFT) has been found.

The second term examination (April 1988)

(a) The sample

Two students left the study, and so the sample size then was 83. Following the same procedure which was employed for the first exam, the distribution of the (HFT) scores of the sample was made and displayed in Figure 22. The classification of the sample into three groups is given in Table 45.

(b) The results

Mean score and standard deviation for the groups' attainment in April exam has been calculated. Table 46 gives the results. While Table 47 shows a clear difference between field dependent students and both field intermediate-independent students. A low correlation has been found between scores of both the exam and the (HFT).

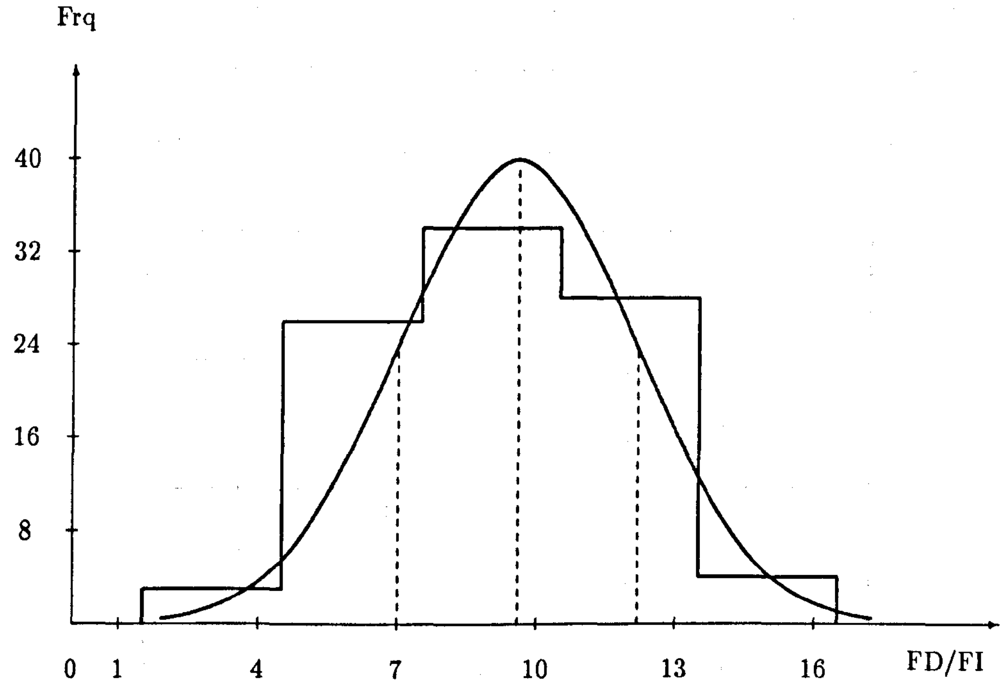

Figure 21

range 13

median 9

mean 9.6

SD 2.6

	FD	FInt	FI
size	19 (22%)	54 (63%)	12 (14%)

Table 42. Classification into FD/FI groups (1st exam "B" class).

group	mean	SD
FD	67.3	14.0
FInt	69.3	15.7
FI	71.9	15.0

Table 43. Mean and SD of FD/FI groups (1st exam "B" class).

	FD	FInt	FI
FD	1		
FInt	NS	1	
FI	NS	NS	1

Table 44. The significant difference in means (1st exam "B" class).

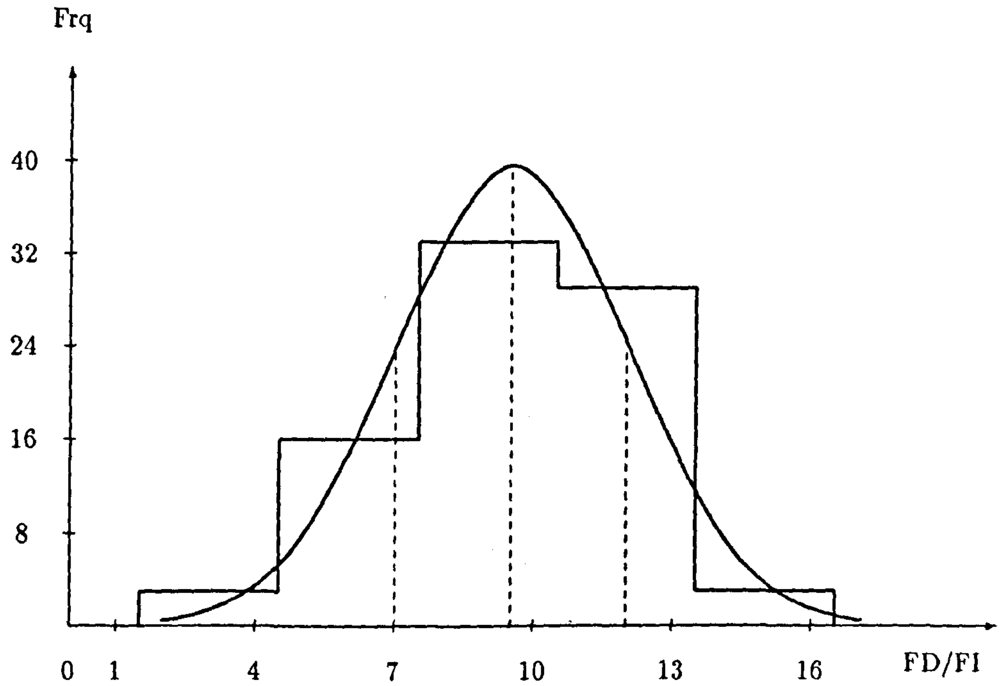

Figure 22

range 13

median 9

mean 9.5

SD 2.5

	FD	FInt	FI
size	19 (23%)	53 (64%)	11 (13%)

Table 45. Classification into FD/FI groups (2nd exam "B" class).

group	mean	SD
FD	46.1	17.6
FInt	59.5	23.6
FI	58.4	21.0

Table 46. Mean and SD of FD/FI groups (2nd exam "B" class).

	FD	FInt	FI
FD	1		
FInt	S (1%)	1	
FI	S (10%)	NS	1

Table 47. The significant difference in means (2nd term "B" class).

Algerian school

The Algerian sample has already been introduced during the analysis of its three terms examinations. The affect of field dependent-independent cognitive style on mathematics performance has again been tested through the comparison between pupils' achievement in their three terms examinations.

The first term examination (December 1988)

(a) The sample

The same sample which attempted the first exam has been tested in the (HFT) too. The size of this sample was 116. The distribution of their scores on the (HFT) is displayed in Figure 23. The pupils then were divided into three groups according to their attainment in the (HFT). Table 48 gives the classification of this sample.

(b) The results

In order to compare the pupils' achievement in the mathematics examination and their scores on the (HFT), Table 49 gives the mean score and standard deviation of each group. While Table 50 shows the significant difference in means within different groups. The difference between means of field dependent-independent groups was found to be significant. This finding then, supports the hypothesis that the field independent pupils perform better than field dependent ones. The Pearson correlation coefficient between pupils' score in both first exam and the (HFT) was found not significant.

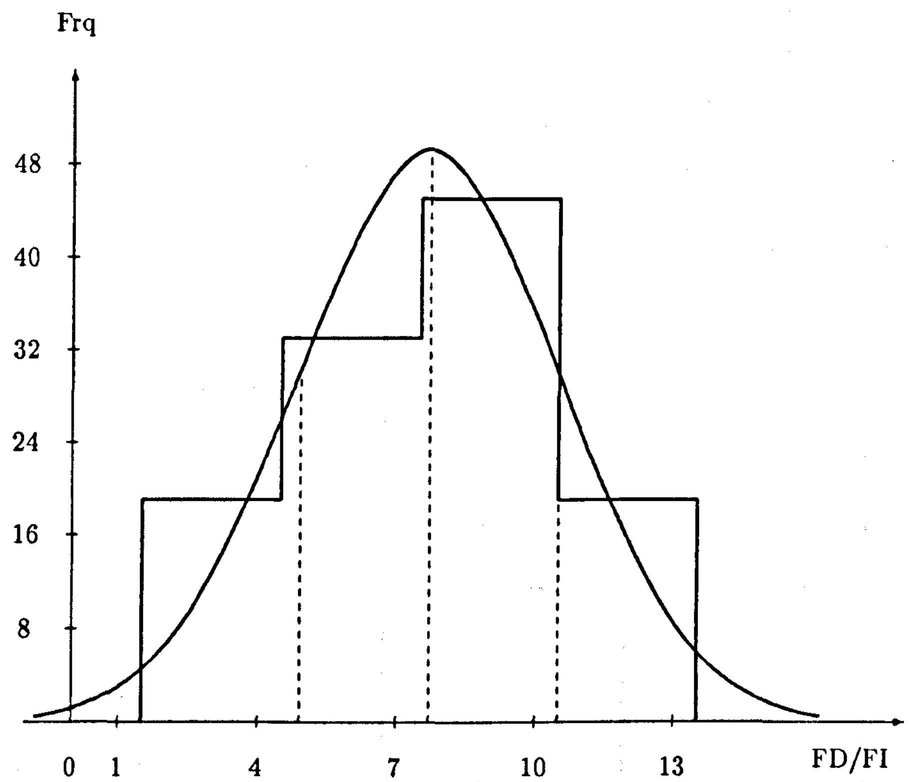

Figure 23

range 11

median 8

mean 7.7

SD 2.8

	FD	FInt	FI
size	33 (28%)	64 (55%)	19 (16%)

Table 48. Classification into FD/FI groups (1st exam Alg. Sch.).

group	mean*	SD
FD	36.2	17.2
FInt	38.8	17.4
FI	45.8	17.6

Table 49. Mean and SD of FD/FI groups (1st exam Alg. Sch.).

* possible score is 80.

	FD	FInt	FI
FD	1		
FInt	NS	1	
FI	S (at 5%)	NS	1

Table 50. The significant difference in means (1st exam Alg. Sch.).

The second term examination (March 1989)

(a) The sample

The sample which attempted the second exam has already been tested in the (HFT). It size was 106. The pupils' scores on the (HFT) has been plotted in Figure 24 and according to their scores, they were divided into three groups as Table 51 shows.

(b) The results

Table 52 gives mean score and standard deviation for each group. Even though performance of field independent pupils was better than field intermediate ones and both were better than field dependent ones, there was no significant difference between them (Table 53). However, a significant correlation ($r = 0.196$, at 5%) between pupils' scores in the exam and the (HFT) has been found. This supports the hypothesis that the lower score is of field dependent pupils, whereas the higher score is of field independent pupils.

The third term examination (May 1989)

(a) The sample

The number of pupils who had scores in both of the third exam and the (HFT) were 97. The distribution of their scores on the (HFT) is shown in Figure 25 and the classification of the sample into three groups is given in Table 54.

(b) The results

Both mean score and standard deviation have been calculated and tabulated (Table 55). Once again the performance of the groups was as expected even though no significant difference in means has been found (Table 56). A correlation of $r = 0.17$ has been found between pupils' scores in the exam and the (HFT).

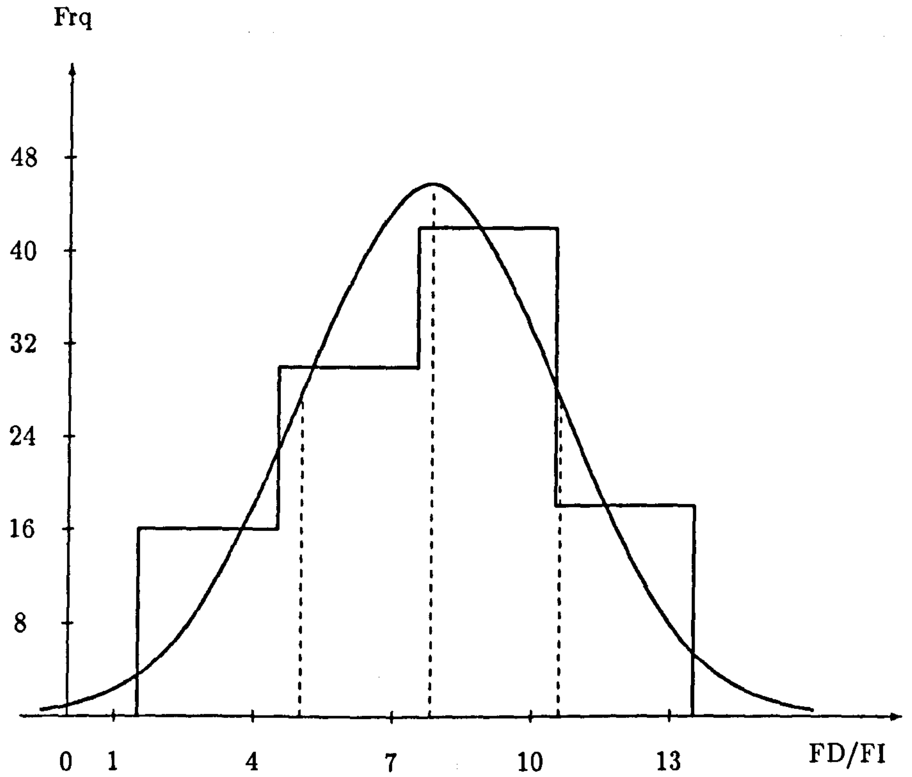

Figure 24

range 11

median 8

mean 7.8

SD 2.8

	FD	FInt	FI
size	28 (26%)	60 (57%)	18 (17%)

Table 51. Classification into FD/FI groups (2nd exam Alg. Sch.).

group	mean*	SD
FD	9.7	4.6
FInt	10.6	4.4
FI	11.2	3.6

Table 52. Mean and SD of FD/FI groups (2nd exam Alg. Sch).

* possible score is 20.

	FD	FInt	FI
FD	1		
FInt	NS	1	
FI	NS	NS	1

Table 53. The significant difference in means (2nd exam Alg. Sch.).

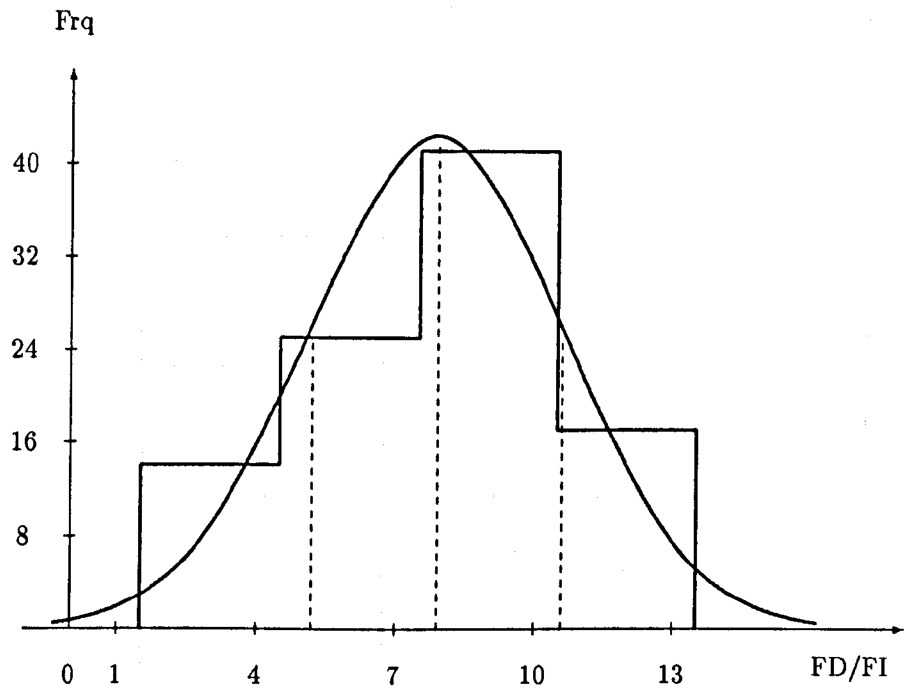

Figure 25

range 11

median 8

mean 7.9

SD 2.7

	FD	FInt	FI
size	24 (25%)	56 (58%)	17 (17%)

Table 54. Classification into FD/FI groups (3rd exam Alg. Sch.).

group	mean*	SD
FD	7.9	2.9
FInt	8.9	3.5
FI	9.2	3.9

Table 55. Mean and SD of FD/FI groups (3rd exam Alg. Sch.).

* possible score is 20.

	FD	FInt	FI
FD	1		
FInt	NS	1	
FI	NS	NS	1

Table 56. The significant difference in means (3rd exam Alg. Sch.).

CONCLUSION

The general conclusion of the results gathered from the University and school samples is that the field dependent-independent cognitive style could influence the performance of the subjects in the mathematics examinations. The comparison between mean scores of field dependent-independent subjects shows that, from seven examinations, the mean score of field independent subjects was higher than of that of field dependent ones in six. However, even though these findings were not significant (just one case), the general and cumulative direction was quite clear.

CHAPTER FIVE

Some Factors Affecting Learning in Mathematics

1. The classification of mental activities involved in learning mathematics

In order to diagnose the difficulty in mathematics, it is necessary to consider the classification of mental activities involved in learning mathematics and then identify their level of thinking. As we already reported (in Gagné's model), Brown [21] suggested four types of mathematical learning: **simple recall, algorithmic learning,** conceptual learning and **problem solving.** But she noticed the difficulty of attempting to categorize tasks according to which type of learning it requires without a knowledge of the previous learning experiences and present conceptual structure of the learner. According to Sharrat's view [1], simple recall is restricted to the memorising of facts, definitions and rules (e.g. multiplication tables, units and simple formulae respectively). Remembering and retrieving concepts have already been discussed during the presentation of memory. Moreover, retention can be fostered with variations in layout of text, placing of certain key elements in boxes and summary notes.

In algorithmic learning (e.g. long multiplication) an individual may act as a computer in terms of recalling and transforming into operations. It was clear that the recall of algorithms becomes easier if they are meaningfully related to the learner's knowledge and many errors result not from failing to learn a particular algorithm but from learning the wrong algorithm.

The lack of understanding in conceptual learning may emerge when structually equivalent symbolic and conceptual tasks are not recognized as being the same. Diagnosis of pupil's existing concepts and the provision of problems based on familiar situations and concrete materials, with conflicts of understanding resolved by discussion, are all helpful. If the learning of concepts has been effective, the pupil should be able to use them to solve problems.

Problem solving implies a process by which the learner combines previously learned elements of knowledge, rules, skills and concepts to provide a solution to a novel situation. The influence of Polya's ideas [118] about how to solve problems was great, since many researchers of mathematical problem solving have tried to improve pupils' ability to solve problems by teaching them Polya's version of heuristic strategies. These are:

(a) understanding the problem,

(b) devising a plan,

(c) carrying out the plan,

(d) looking back.

The first step may involve substeps such as drawing a diagram, choosing appropriate notation, whether the information provided is sufficient and whether it incorporates any redundancy.

The middle two steps are difficult particularly devising a plan since creativity and insight might be required.

The last step involves the final checking and whether the result may be generalised and whether alternative solutions may exist.

Sleet et al. [119] suggested the following stages for solving problems: representing or defining the problem, devising a plan and solving the problem and checking and reviewing. The similarity with Polya's ones is obvious. Sleet confirmed that possession of all the prerequisite skills is not necessarily sufficient to enable a student to solve a problem. This may be due to the load on a student's working memory being greater when the subproblem has to be extracted from a main problem than when it is separated from it. As a result of this, developing a plan to solve a problem is very important. He noted a lack of students' confidence in processing in a problem.

It was noted that [120] in solving problems, the better mathematics students focus on the mathematical structure of the problem. They single out the basic mathematical relations in a problem which are essential for its solution and ignore the superfluous information in the problem statement. While less capable students attend to irrelevant information in a problem and do not isolate the critical mathematical relations. It was also noted that there were different kinds of mathematical ability: some pupils have an "analytic" mind and preferred to think in verbal, logical ways. Other pupils have a "geometric" mind and liked a visual or pictorial approach. Some other pupils have a "harmonic" mind and are able to combine characteristics of both the analytic and the geometric.

2. Understanding

A psychological theory of complex behaviour is an asset if it can close - or at least reduce - the gap between students' aspirations and difficulties in learning caused by the education system (i.e. the curricula, materials, etc.). The psychological theories of complex behaviour can be classified [121] into three categories called descriptive, explanatory and predictive. Predictive theories are the most important since we wish to predict success and failure in learning situations. An example of this category, is the Model of Thinking/Memory Capacity proposed by Johnstone and El-Banna [109]. This model (which already has been presented) is based on information processing theory which enhances our understanding of learning and thinking both in classroom and out of it [122].

It is interesting to find out why pupils have difficulty in learning. The answer might be found through attempting to understand the understanding itself. Dobson [123] stated that an idea is understood if it can be used by the learner. It is clear that the value of this view comes from its applicability for assessing what pupils understood. The meaning of understanding has three different aspects which are [124]:

(i) to understand some things is to see them;

(ii) understanding some things as being the ability to construct a useful mathematical model of it;

(iii) to understand some things is to know how to find (or construct) a plausible schema which allows one to assimilate it to what one already knows.

The first aspect is the every day view which encourages visual imagery which is very important not just for some physical phenomena but also for some classroom situations.

The value of the second view [125] may come not just from its applicability in mathematics, since any mathematical system or structure is a potential model, but also in other subjects. The reason for this may due to the characteristic of mathematical modelling as being:

(a) realistic (since the unrealistic result is caused, not by the internal logic of the model, but by the assumptions underlying the model);

(b) **predictive** (e.g. the prediction of hypothetical or real life situations);

(c) **generative** (which leads to the concept of a hierarchy of relationships and the idea of a mathematical system).

The third aspect derives from schema theory. The schema is a very important tool since any scientific theory can be regarded as a type of it, as a mental representation used to make sense of some part of nature [124]. The most important value of this theory is to explain why pupils do not understand certain material. Three reasons for this misunderstanding are listed [126]:

(1) a person does not have an appropriate schema which can assimilate certain material;

(2) a person knows an appropriate schema but a given situation does not elicit it;

(3) students have a competing schema that they use to comprehend material.

Although the above reasons seem to be resonable for failure to understand a phenomenon some researchers (e.g. Driver [127]) have gone for the third one which is known under the names of **"alternative frameworks"** and **"misconceptions"** since children often strongly hold alternative schemata that they use to make sense of the world and the science curriculum [124].

The careful analysis of the "misunderstood" situations listed above, may lead to the following:

(i) the lack of an appropriate schema may come from teachers who may not sufficiently emphasize the learning of specific schemata [124];

(ii) a possible reason for the second situation is that some areas of a subject matter are too difficult to handle since the degree of familiarity and abstraction of mode and idea play a role [128];

(iii) a third case is when students already hold a competing schema, it is possible to focus on how a learning process occurs in a typical classroom and out of it and then, the emphasis goes to the diagnostic teaching of strategies as a way for dealing with this situation.

We already indicated the necessity of assessing what pupils understood but "**testing**" may not be a helpful tool if we do not define the exact goal of what we are assessing. Is it possible to outline the above factors (i.e. the way of teaching, learning, testing and the nature of the subject matter) as the most important issues which affect the learning difficulties? This opinion was also suggested by Johnstone [76] who noted that the **nature of science**, the **traditional method of teaching** and the **way of learning** are three possibilities at least by which these difficulties have arisen.

Because there is a limitation on what we can understand, the idea of "**level of understanding**" was introduced by Skemp [129] but a lot of work has been done by other researchers to illustrate and extend it. Skemp [130] suggested three levels of understanding called "**instrumental**", "**rational**" and "**formal or logical**" understanding and defined as follows:

(i) the first is the ability to apply an appropriate remembered rule to the solution of a problem without knowing why the rule works;

(ii) the second is the ability to deduce specific rules or procedures from more general mathematics relationships;

(iii) the third is the ability to connect mathematical ideas and to combine these ideas into chains of logical reasoning.

According to Skemp, the goal of instrumental learning is to be able to give right answers, whereas the goal of rational understanding is the ability to specify methods for particular problems and the goal of logical understanding is the construction of chains of logical reasoning to produce what we call demonstrations or proofs.

A mis-match can occur between pupil and material (question for example) if the pupil's conception of understanding is instrumental while the aim of the question is rational (or logical) understanding. It was noted that [1] there was difficulty in making valid inferences about whether a person understands rationally or instrumentally from his written work, but talking with him is the best way to find out even though this is difficult to achieve in large classes.

To sum up, the idea of level of understanding is very important for the diagnosis of the degree of difficulties encountered by pupils. While, understanding itself is a basic goal of education both educators and educational psychologists hope that pupils benefit from their work, the purpose of which is to reduce learning difficulties.

3. Difficulties in mathematics

According to Macnab and Cummine [61], the causes of learning difficulties in mathematics stem from the nature of the subject itself, its thought processes and its symbolism. The following headings may illustrate such difficulties:

1. the abstract nature of the concepts involved,

2. the complexity of mathematics,

3. formal notation,

4. formal algorithms,

5. the concepts and use of variables,

6. spatial concepts and geometric thinking.

In teaching, one should accept that all mathematical ideas are complex in order to overcome any problem which may arise. This suggests the simplicity through abstraction and analogy.

To reduce learning difficulties related to the content hierarchy of mathematics, revision and looking ahead may be required. It was noted that, at all levels of mathematical competence, some ability in logical thought is necessary in order to understand the formal deductive side of mathematics and the process aspect of proof.

Formal notation in mathematics can cause considerable confusion in the minds of many pupils and the ability to use appropriate notation effectively takes time and experience to develop. The following principles were suggested [61]:

(i) the meaning of mathematical symbols and notation should be precise;

(ii) the problem caused by visual appearance should be taken into account;

(iii) associate manipulation with meaning, develop in pupils the ability to carry out manipulation correctly without continual recourse to semantic interpretation;

(iv) be aware of the anomalous side of mathematical notation.

The negative side of algorithms may be that they interfere with reasoning abilities and mathematical thinking. While the positive aspect of them is that they enable complex processes to be carried out by a very much simpler technique.

Variables can cause considerable confusion in learning for three main reasons:

(i) they may be introduced to pupils in contexts in which their purpose is not obvious;

(ii) they may be introduced in contexts where the notion of variability is not evident;

(iii) distinctions may not have been made between variables in the ordinary sense and descriptive or bound variables.

Learning difficulties in geometry can arise because:

(i) geometrical truths have to be distinguished from accidently irrelevant features of particular diagrams;

(ii) observation must be distinguished from logical consequence;

(iii) exact theoretical calculation must be distinguished from particular measurement;

(iv) reflective insight is necessary to perceive implicit aspects of geometrical diagrams;

(v) it is necessary to be able to comprehend three-dimensional objects and their properties through two-dimensional representation.

The difficulty in mathematics is not only caused by subject matter, but also by other factors. A study carried out by Pollitt et al. [128] found that the difficulty appeared to be affected by such aspects as the need for a strategy, the level of abstraction of the mathematical language used, the provision of concrete reference material (diagrams), the number of reasoning or calculation steps involved and the demand of comprehension rather than knowledge.

This study suggested that the necessity for comprehension or analysis is a problem and the need to seek out a strategy for answering the question makes questions considerably more difficult. In their conclusion, they divided the difficulty into three categories:

subject or concept difficulty, where one particular concept may be more difficult than another and usually the reasons for difficulty are the degree of familiarity and abstraction of mode and idea;

process difficulty, where a particular operation or sequence of operations demands manipulation of data at a level beyond the straightforward recall of specific learned items;

question or stimulus difficulty, where the guidance given to candidates in directing their attention to a particular response, or the support given in terms of additional information or data, is either minimal or non-specific.

Students' mistakes in problem solving have been discussed and listed [131]:

(a) many of the difficulties encountered during problem solving are due to the use of an incorrect method;

(b) "how and where does one start" seems to be one of the major problems that students encounter during problem solving;

(c) a major difficulty may occur when students do not define the goal or do not clarify the problem or do not proceed step by step.

4. Language in mathematics

Studies of the effect of language on students' performance on multiple choice tests provide further evidence of how the representation of a problem can affect students' ability to solve it.

According to Cassels and Johnstone [84], wordy questions with embedded clauses seem to be more difficult than short questions written in short sentences. They also stated that the removal of negatives in general seems to improve the performance.

In order to find out how language functions during problem solving, Goldin [132] listed four kinds of language which he thought appeared. These are:

1. language of problem statement,
2. non-verbal language (diagram etc.),
3. notational language,
4. planning language.

The planning language is used to talk about the other three kinds, to review and make plans about what to try. The role of language in the problem solving process may appear in the following classification of errors [133]: reading ability, comprehension, transformation, process skills and encoding. While the two other types of errors which are motivation and carelessness can occur at any stage of the above. It was noted that the question form is also another source of error.

It was noted that [134] mathematics language facilitates thinking by complementing ordinary language and it also suggests solutions to problems. A perfect language probably would embody the principle of one word for one idea. Ashlock went further when he stated that [8] the essential difficulties appear to have nothing overtly to do with mathematics since the abstraction process is made difficult by the unfamiliar context rather than the mathematical concepts.

The work which has been done by Shuard and Rothery [135] contains different forms of language such as wording (verbal), symbolic and graphic forms. While the visual appearance is another important factor which has been discussed too.

Wording form

The terms Ordinary English (OE) and mathematical English (ME) were introduced [136] to help stress the special nature of written mathematics. The differences between (OE) and (ME) make it impossible to apply standard readability formulae to mathematics books. Wording can be divided into three categories [135]:

(i) words which have the same meaning in (ME) as in (OE) (e.g. because);

(ii) words which have a meaning only in (ME) (e.g. hypotenuse);

(iii) words which occur in both (OE) and (ME) but which have a different meaning in (ME) from their meaning in (OE) (e.g. difference).

Except the first category in which the words are familiar to the children, the other two caused a lot of difficulty since they are rare in the child's experience or they have variety of meaning. According to Johnstone [116], if a word we use is familiar to a student, but his meaning and ours differ, we are unlikely to generate the learning we want. In short, the familiarity of words is perhaps a better measure of their difficulty.

Symbolic form

It was reported that [135] there were two different coding systems used in mathematical texts, that which in used for words and is based in sound and that which is used for signs and symbols and which is based on pictorial form. When children learn symbols, they need to link together three things: an idea, some words which correspond to that idea and the symbol. Many difficulties which children encounter arise from the complicated nature of the interrelationships between these three.

Graphic form

Some examples of graphic language forms used in mathematics are [135] tables, graphs, diagrams, plans and maps and pictorial illustrations. Illustrative matter can be classified in terms of how importantly it is related to the prose text. Three levels of importance can be found in the illustrations of children's mathematical texts: decorative, related but non-essential and essential.

The visual appearance of the text

The appearance of a page of a text depends [135] upon factors which are not usually consciously considered by the teacher or the pupil. It seems likely that the visual appearance of a good page of a text will be:

easy for the reader to find his way about;

pleasing to look at.

These qualities can be achieved through a careful choice of the layout of the page, the type style used in printing and the use of colour.

To sum up, it was found that [116] an unfamiliar word, or a known word used in an unfamiliar way, takes up valuable working space. A question presented in a negative form takes up more working space than in a positive form. Double negatives can take up three or four working spaces, and triple negatives blow the system completely!

The difficulties involved in reading mathematics may be due [61] to syntactic complexity of the English used, use of technical vocabulary, the mathematical notation used and inability to relate the mathematics to the context. However, improving the text, the teacher's use of the text and the reading ability of the reader are all recommended for ameliorating language problems [135].

CHAPTER SIX

The study of Paper I

Introduction

In order to find out some factors which might affect the difficulty of mathematics questions and clarify the idea of demand, an investigation was carried out in some schools presenting for the Scottish Examination Board (SEB) Higher Grade examination. The preliminary examination of five schools in 1989 and the Scottish Certificate of Education (SCE) Examination in 1988 have been analysed. In the Higher Grade examination at that time, two papers were involved to assess pupils' ability. Paper I, of 1¾ hours duration, contained forty multiple-choice items each with five responses. Paper II, of 2½ hours duration, contained thirteen more traditional questions. Four ability levels were tested by Paper I [137]: knowledge, comprehension, application and analysis/evaluation. Paper II assessed the competence of the candidates to perform manipulation, to reproduce set work and to sustain logical thought. In other words, this paper provided a balance in terms of coverage of syllabus content and assessment of the following abilities: communication, systematic problem solving, data analysis and creativity [46].

Past multiple-choice items were often used as revision tests within the classroom, while the traditional questions were more often given as homework or used in school examinations. In terms of diagnosis, if pupils succeed in their answers, we may assume that they have attained the desired level of understanding. But if their answers are wrong, is it possible to deduce that pupils have failed to reach the desired level of understanding? If this is the case, can we identify where the failure has occurred, what factor or factors are have been involved and how do these factors affect the demand of questions? To reach these goals, the preliminary examination of five schools (both Papers I and II) and the SCE Higher Grade examination (Paper II) have been analysed. The questions from both papers can be found in Appendix 9.

The analysis of Paper I

Note that, Paper I of School 5 - contrary to that of other schools - contains twenty - four short traditional items. The sample size was 252 pupils in five separate schools. The distribution of sample size according to schools is shown in Table 57. The results for each school have been analysed.

1. School 1

The distribution of test scores is given in Table 58, while Figure 26 illustrates these results. The distribution is, in general, normal as indicated by the closeness of fit to the normal curve with the same mean and standard deviation as the original distribution. Note that the top 27% of the sample scored from 26 to 31 (out of 40), whereas the bottom 27% ranged from 9 to 18.

The Facility Value (FV) is a useful signal statistic for any given test item. As a general principle adopted in this study, an item is described as "easy", "average", or "difficult" depending on whether its FV is greater than 0.6, between 0.4 and 0.6, or less than 0.4 respectively. The Point Biserial Correlation Coefficient (PBCC) is another useful statistic for measuring an item's discrimination: in this study an item is described as "high", "acceptable", or "poor" in discrimination depending on whether its PBCC is greater than 0.6, between 0.4 and 0.6, or less than 0.4. Both FV and PBCC have been calculated for each item and tabulated with the evaluation (EV) (Table 59). Note that the percentage of "easy", "average" and "difficult" items is 43, 32 and 25 respectively. Ten items are classified as "difficult", two of them are acceptable while the others are "poor" (including the "null" and "negative" discrimination items). The latter items certainly caused difficulties for pupils arising from mathematical factors (period, collinarity, mapping, etc.) and other factors (formulation, unclear diagram, abstract form, requirement to show understanding, etc.). Note that half of these items (i.e. 4 items) were multiple-completion types. This type may need more effort than the multiple-choice type.

school	size
School 1	34
School 2	50
School 3	96
School 4	23
School 5	49
total	252

Table 57. The sample size.

Frq	Mark	
2	31	the top 27% of the sample (*N* = 9)
3	28	
2	27	
2	26	
2	25	
4	24	
3	23	
1	22	
3	21	
2	20	
2	18	the bottom 27% of the sample (*N* = 9)
2	17	
3	16	
2	13	
1	09	

Table 58. The distribution of the test scores of School 1.

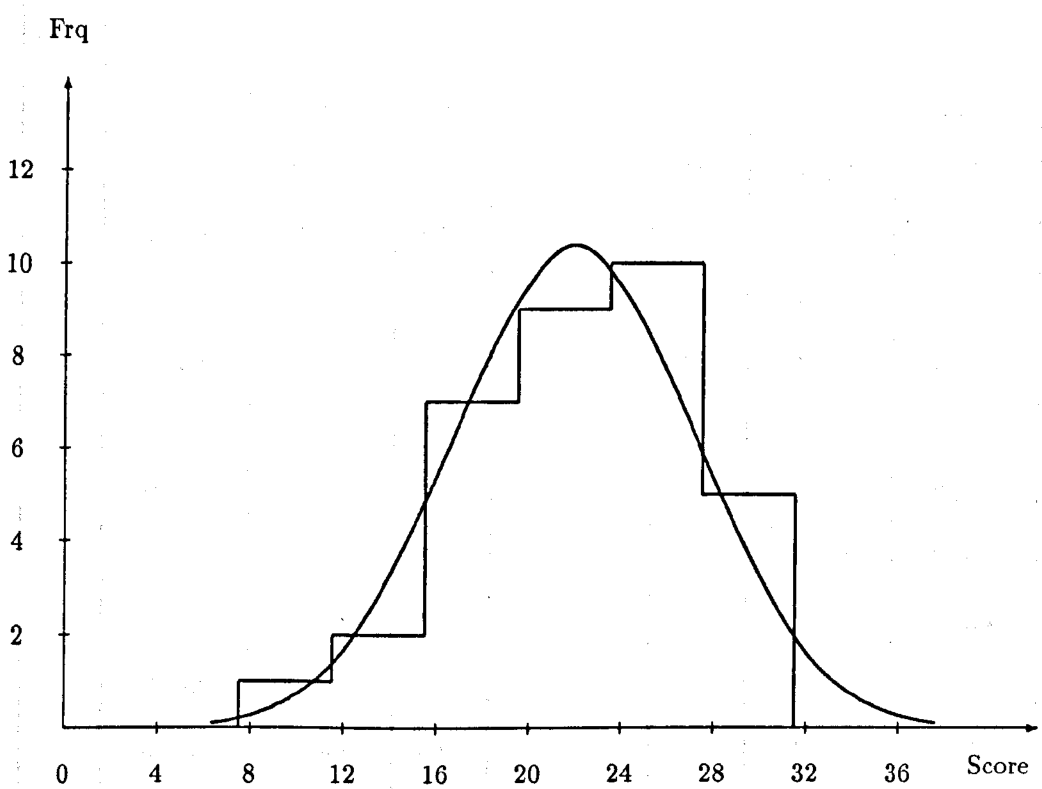

Figure 26

range 22

median 23

mean 21.9

SD 5.2

item	FV	PBCC	EV
1	0.97	0.20	EP
2	0.79	0.47	EA
3	0.74	0.31	EP
4	0.88	0.52	EA
5	0.62	0.18	EP
6	0.88	0.36	EP
7	0.76	0.43	EA
8	0.82	0.63	EH
9	0.62	0.54	EA
10	0.62	0.42	EA
11	0.88	0.57	EA
12	0.56	0.54	AA
13	0.09	0.18	DP
14	0.74	0.46	EA
15	0.74	0.17	EP
16	0.06	0.00	DN
17	0.44	0.19	AP
18	0.56	0.37	AP
19	0.50	0.33	AP
20	0.47	0.32	AP
21	0.50	0.41	AA
22	0.91	-0.01	EP
23	0.74	0.43	EA
24	0.47	0.46	AA
25	0.47	0.47	AA
26	0.21	0.09	DP
27	0.32	0.46	DA
28	0.56	0.34	AP
29	0.44	0.36	AP
30	0.97	0.20	EP
31	0.76	0.42	EA
32	0.53	0.54	AA
33	0.32	0.42	DA
34	0.35	-0.35	DP
35	0.15	0.13	DP
36	0.44	0.11	AP
37	0.47	0.15	AP
38	0.15	-0.26	DP
39	0.15	0.05	DP
40	0.26	-0.03	DP

Table 59. The FV and PBCC for each item of School 1.

Key: EP easy-poor, EH easy-high, AA average-acceptable, DN difficult-null, etc.

2. School 2

The test scores are given in Table 60, while Figure 27 shows the distribution of these marks. Note that the **top 27%** of the sample scored from 26 to 34, whereas the **bottom 27%** ranged from 13 to 19. Both FV and PBCC have been calculated for each item. This is shown in Table 61. It was found that the percentage of "easy" and "average" items was 40 in each case, while the percentage of "difficult" ones was 20. Among the last category, there were two items which were acceptable, but the remaining six were poor (including a negative discrimination item). The difficulty of these items, in general, comes from the subject matter (trigonometic formulae, numeration, limit, etc.) and from other factors such as the need for understanding, and a diagram holding too much information.

Frq	Mark
2	34
2	33
1	32
1	31
1	30
3	28
3	27
3	26
4	25
5	24
4	23
2	22
1	21
2	20
4	19
7	18
1	17
1	16
1	14
2	13

(Marks 34–26:) ⟵ the top 27% of the sample ($N = 14$)

(Marks 19–13:) ⟵ the bottom 27% of the sample ($N = 14$)

Table 60. The distribution of the test scores of School 2.

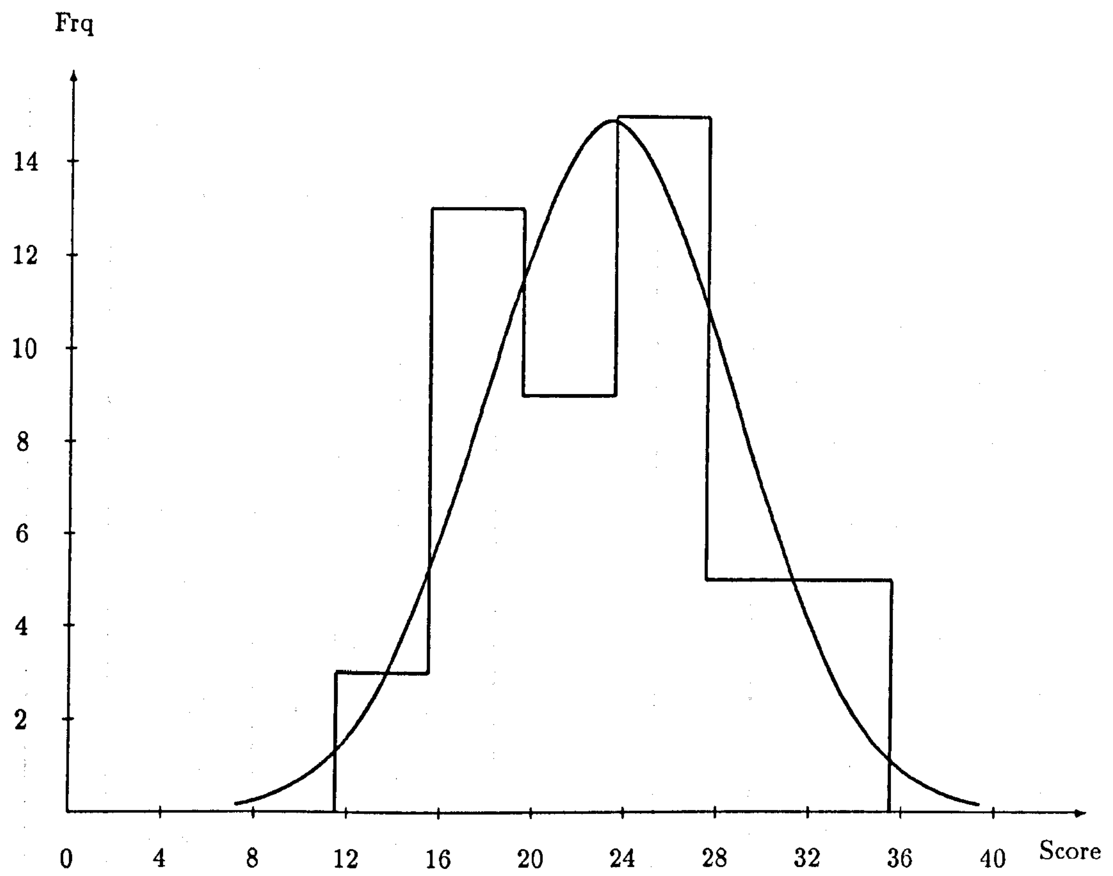

Figure 27

range 21

median 23.5

mean 23.2

SD 5.4

item	FV	PBCC	EV
1	0.86	0.16	EP
2	0.94	0.09	EP
3	0.72	0.22	EP
4	0.78	0.31	EP
5	0.60	0.40	AA
6	0.88	0.41	EA
7	0.88	0.35	EP
8	0.58	0.39	AP
9	0.10	0.40	DA
10	0.48	0.31	AP
11	0.70	0.32	EP
12	0.56	0.01	AP
13	0.82	0.36	EP
14	0.88	0.10	EP
15	0.66	0.32	EP
16	0.56	0.27	AP
17	0.26	-0.02	DP
18	0.74	0.57	EA
19	0.68	0.25	EP
20	0.50	0.57	AA
21	1.00	0.00	EN
22	0.54	0.56	AA
23	0.46	0.18	AP
24	0.34	0.31	DP
25	0.42	0.58	AA
26	0.26	0.26	DP
27	0.74	0.40	EA
28	0.74	0.28	EP
29	0.38	0.52	DA
30	0.48	0.24	AP
31	0.42	0.33	AP
32	0.34	0.21	DP
33	0.92	0.15	EP
34	0.58	0.27	AP
35	0.26	0.17	DP
36	0.10	0.19	DP
37	0.42	0.19	AP
38	0.58	0.45	AA
39	0.56	0.40	AA
40	0.42	0.42	AA

Table 61. The FV and PBCC for each item of School 2.

Key: EP easy-poor, AA average-acceptable, EN easy-null, etc.

3. School 3

Table 62 shows the distribution of the test scores, which is illustrated in Figure 28. Once again, note that this distribution is, in general, normal. The top 27% of the sample scored from 27 to 39, while the bottom 27% ranged from 9 to 19. For each item, FV and PBCC have been calculated, the results are shown in Table 63. It was found that the percentage of "easy", "average" and "difficult" items was 45, 40 and 15 respectively.

According to Table 63, there are six difficult items, three of them are acceptable, but the other three are poor. The difficulty of the first group may be due to mathematical factors (trigonometic formulae, length of a vector) or to graphic language (identification of a property from a set of graphs). The difficulty of the second group may arise from trigonometric formulae, irrelevant and abstract formulation, and the need to show understanding.

Note that, of the two items of multiple-completion type, one was found difficult.

Frq	Mark
1	39
2	38
1	37
1	36
1	35
2	34
2	33
2	32
3	31
5	30
2	29
3	28
5	27
6	26
3	25
5	24
6	23
7	22

the top 27% of the sample (N = 26) ← (marks 39–27)

5	21
3	20
8	19
3	18
8	17
3	16
4	15
1	13
1	12
1	11
1	10
1	09

the bottom 27% of the sample (N = 26) ← (marks 19–09)

Table 62. The distribution of the test scores of School 3.

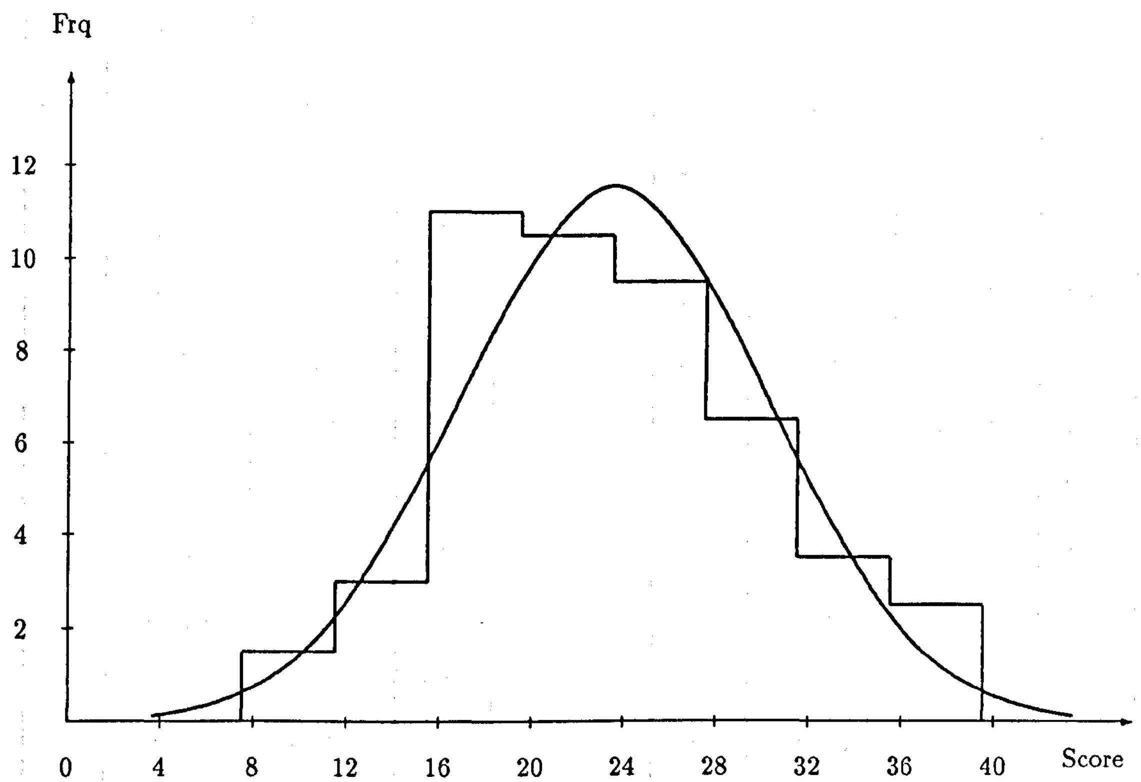

Figure 28

range 30

median 23

mean 23.5

SD 6.6

item	FV	PBCC	EV
1	0.73	0.36	EP
2	0.77	0.41	EA
3	0.58	0.20	AP
4	0.50	0.33	AP
5	0.79	0.30	EP
6	0.66	0.51	EA
7	0.56	0.57	AA
8	0.83	0.40	EA
9	0.66	0.38	EP
10	0.83	0.34	EP
11	0.94	0.36	EP
12	0.52	0.52	AA
13	0.48	0.37	AP
14	0.34	0.29	DP
15	0.52	0.43	AA
16	0.51	0.48	AA
17	0.48	0.12	AP
18	0.50	0.37	AP
19	0.41	0.32	AP
20	0.79	0.32	EP
21	0.89	0.18	EP
22	0.61	0.50	EA
23	0.82	0.21	EP
24	0.50	0.30	AP
25	0.31	0.25	DP
26	0.78	0.44	EA
27	0.41	0.24	AP
28	0.65	0.25	EP
29	0.57	0.38	AP
30	0.54	0.38	AP
31	0.71	0.31	EP
32	0.36	0.45	DA
33	0.65	0.44	EA
34	0.37	0.45	DA
35	0.68	0.45	EA
36	0.58	0.42	AA
37	0.42	0.30	AP
38	0.35	0.45	DA
39	0.60	0.26	EP
40	0.26	0.27	DP

Table 63. The FV and PBCC for each item of School 3.

Key: EP easy-poor, AA average-acceptable, DA difficult-acceptable, etc.

4. School 4

The same method which was applied to above schools has been adopted here. Table 64 gives the distribution of the test scores, whereas Figure 29 illustrates the results. This table shows that the top 27% of the sample scored from 29 to 36, while the bottom 27% went from 13 to 19. Both FV and PBCC for each item has been calculated (Table 65) and the percentage of "easy", "average" and "difficult" items was found to be 48, 25 and 27 respectively.

Note that two items (i.e. items 13 and 38*) are omitted from the following analysis since there was an error in the second and possibly also in the first. Five items (i.e. items 33, 34, 36, 38 and 40) were different for the two classes in the sample. Of the eleven items which were found to be difficult, one was high in discrimination, six were acceptable and four were poor (including the one which has negative discrimination). The main difficulties of these items may arise from parallelism and perpendicularity of vectors, limit, inverse of a function, trigonometric formulae, scalar product, notation, negative form and the need to find a strategy.

Frq	Mark
1	36
2	32
1	31
1	30
1	29
2	28
1	27
1	26
1	24
2	22
2	21
3	19
2	18
1	15
2	13

(Marks 36–29 bracketed:) ⟵ the top of 27% of the sample (N = 6)

(Marks 19–13 bracketed:) ⟵ the bottom of 27% of the sample (N = 6)

Table 64. The distribution of the test scores of School 4.

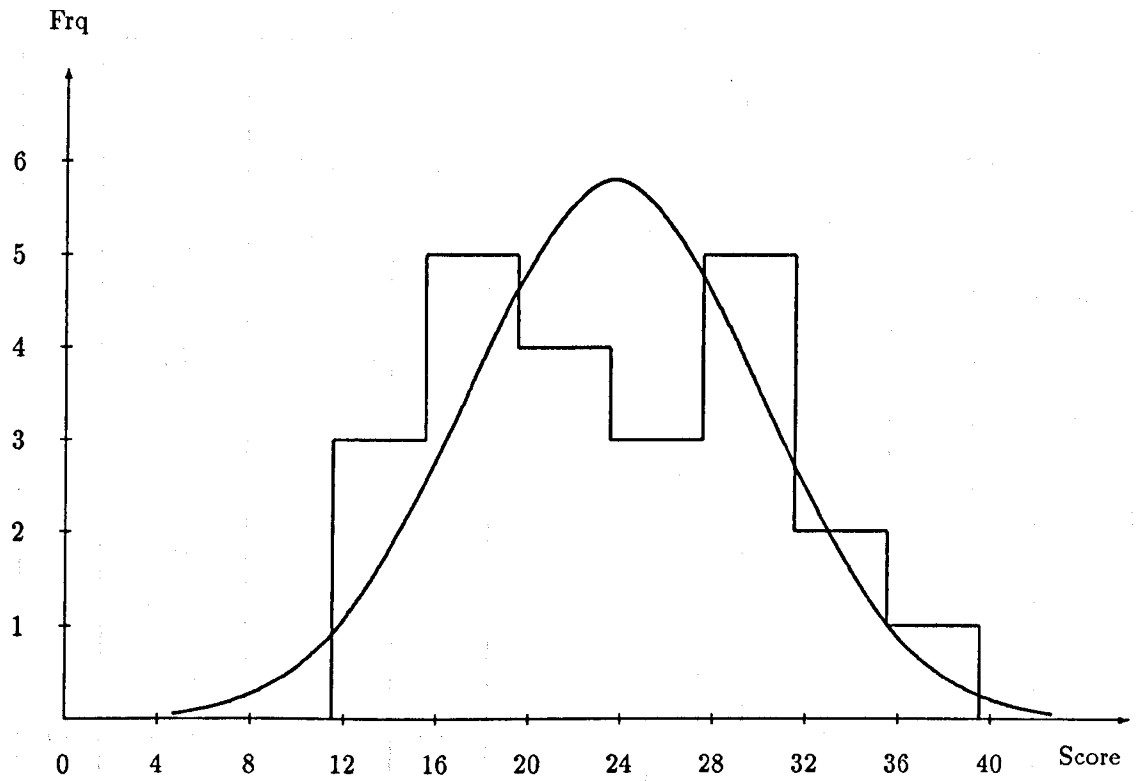

Figure 29

range23

median 22

mean 23.6

SD 6.3

item	FV	PBCC	EV	
1	0.74	0.47	EA	
2	0.78	0.07	EP	
3	0.61	0.31	EP	
4	0.91	0.23	EP	
5	0.52	0.59	AA	
6	0.78	0.61	EH	
7	0.57	0.57	AA	
8	0.65	0.27	EP	
9	0.70	0.58	EA	
10	0.48	0.42	AA	
11	0.52	0.61	AH	
12	0.74	0.23	EP	
13	0.35	-0.07	DP	← omit
14	0.83	0.24	EP	
15	0.78	0.20	EP	
16	0.57	0.51	AA	
17	0.96	0.10	EP	
18	0.61	0.18	EP	
19	0.35	0.45	DA	
20	0.70	0.49	EA	
21	0.83	0.51	EA	
22	0.39	0.41	DA	
23	0.43	0.17	AP	
24	0.30	0.19	DP	
25	0.78	0.38	EP	
26	0.39	-0.09	DP	
27	0.22	0.49	DA	
28	0.70	0.58	EA	
29	0.70	0.51	EA	
30	0.39	0.49	DA	
31	0.61	0.58	EA	
32	0.52	0.62	AH	
33*	0.25	0.73	DH	
33**	0.82	0.35	EP	
34*	0.83	0.15	EP	
34**	0.55	0.52	AA	
35	0.57	0.22	AP	
36*	0.33	0.24	DP	
36**	0.55	0.52	AA	
37	0.43	0.23	AP	
38*	1.00	0.00	EN	omit →
38**	0.36	0.51	DA	
39	0.39	0.31	DP	
40*	0.33	0.49	DA	
40**	0.64	0.71	EH	

Table 65. The FV and PBCC for each item of School 4. Key:* first class only, ** second class only, DH difficult-high, AA average-acceptable, etc.

5. School 5

As we reported earlier, Paper I of this school - in contrast to all other schools - contains twenty-four short traditional items. Therefore in analysing this paper, the **Discrimination Index** (DI) (rather than the PBCC) and the **Reliability Coefficient** (r) (for the test as a whole) are used.

The test scores have been arranged in a rank order (Table 66) and the distribution is illustrated in figure 30. The **top 27%** of the sample scored from 55 to 65 (out of 72), while the **bottom 27%** went from 7 to 36. The calculation of DI needs a new sample which is composed of the top and the bottom of the original sample (in our case both the top and the bottom were 27% of the original sample). FV_1 and FV_2 refer to the facility values of the original and new samples respectively. FV_1, FV_2 and DI have been calculated for each item and the reliability coefficient was found (Table 67). It was found that there was no significant difference in means of facility values of the two samples, and a significant correlation between the facility values of the two samples. Therefore, the new sample can been considered to have the same characteristics as the original one.

The percentage of "easy", "average" and "difficult" items is 54, 29 and 17 respectively.

Frq	Mark
2	65
1	64
3	63
1	61
1	60
1	59
1	58
1	57
4	55
2	54
1	52
3	51
1	50
2	49
1	48
1	47
3	45
2	44
1	43
1	41
1	40
1	39
...	...
1	37
1	36
1	35
2	34
1	33
1	28
1	26
1	24
2	19
1	16
1	11
1	07

Marks 65–55: ← the top 27% of the sample (*N* = 13)

Marks 36–07: ← the bottom 27% of the sample (*N* = 13)

Table 66. The distribution of the test scores of School 5.

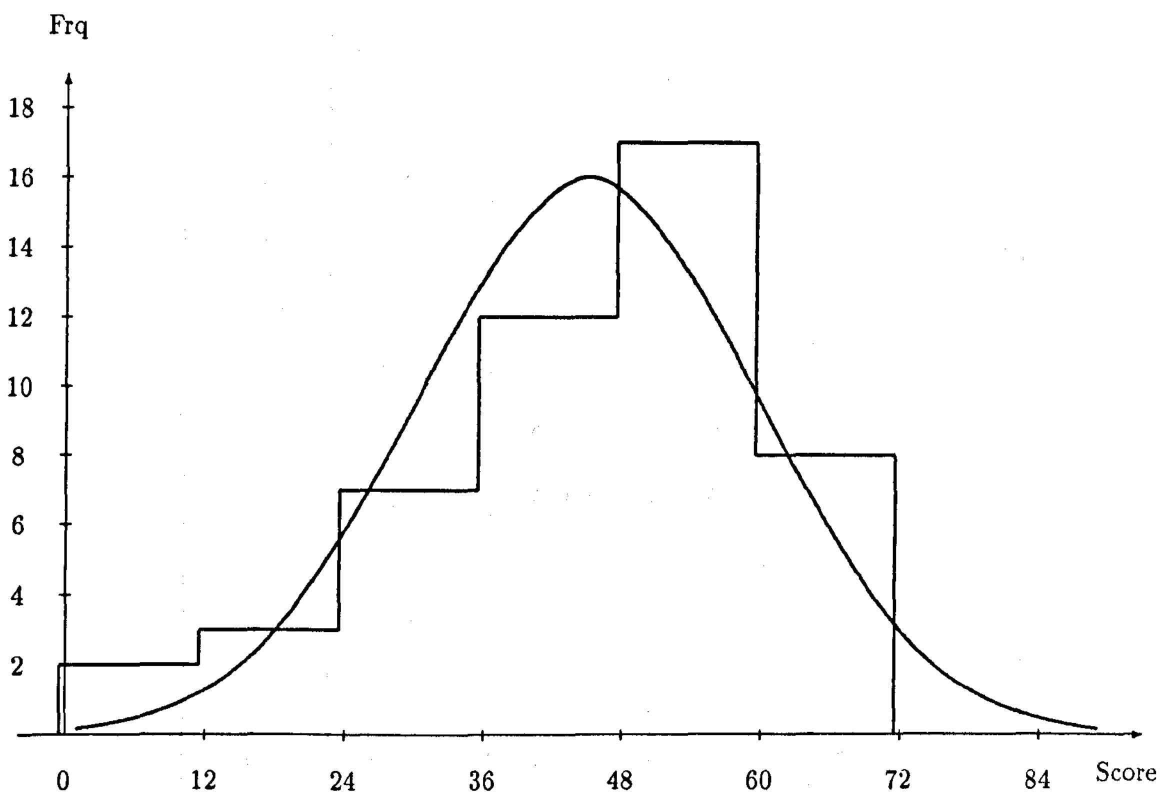

Figure 30

range 58

median 48

mean 44.9

SD 14.7

item	FV_1	FV_2	DI	evaluation
1	0.76	0.81	0.38	easy
2	0.63	0.54	0.77	acceptable
3	0.78	0.73	0.38	easy
4	0.63	0.62	0.77	acceptable
5	0.63	0.62	0.46	acceptable
6	0.59	0.58	0.85	acceptable
7	0.71	0.69	0.15	rejected
8	0.71	0.65	0.54	acceptable
9	0.71	0.62	0.77	acceptable
10	0.55	0.54	0.92	acceptable
11	0.80	0.69	0.46	acceptable
12	0.63	0.73	0.54	acceptable
13	0.29	0.39	0.31	difficult
14	0.37	0.27	0.38	difficult
15	0.55	0.54	0.92	acceptable
16	0.12	0.19	0.23	difficult
17	0.67	0.54	0.77	acceptable
18	0.65	0.69	0.46	acceptable
19	0.43	0.46	0.62	acceptable
20	0.55	0.46	0.62	acceptable
21	0.45	0.50	0.38	improvable
22	0.55	0.54	0.92	acceptable
23	0.71	0.73	0.54	acceptable
24	0.14	0.19	0.38	difficult

Table 67. The FV and DI for each item of School 5 ($r = 0.84$).

In order to analyse the test items in terms of "good" or "poor" items, the two factors of FV and DI have been used together and the evaluation of each item has been made according to Macintosh and Morrison's criteria [138] which is illustrated in Table 68. In this table, the first row's items are in general acceptable, the second row's items are reasonable, but can usually be improved, the third row's items are marginal and usually require substantial revision, and finally the fourth row's items are completely unsuitable.

The evaluation of the test items according to the above criteria has already been made (Table 67). It was found that four items were classified as difficult, the main reasons for the difficulties could be caused by the relationship between two variables, scalar product, trigonometric formulae and the increasing/decreasing properties of a function.

Study of the difficult items

In order to identify the real causes of the difficulties in multiple-choice items, the analysis of all items where FVs were below 0.4 has been made for all schools (except School 5 which had no such type). Table 69 gives the classification of the 35 difficult items according to their contents. For each item, the percentage choosing the right answer and the percentages choosing the main wrong answers have been calculated and an attempt to explain the errors has been made. This can be found in nine tables (from Table 70 to Table 78) and the items can be found in Appendix 9. The following examples illustrate such analysis.

(i) The first item in Table 70 is a multiple-completion type. The similarity in form has been used by pupils to identify a second function after they succeeded in choosing the first (the first two types of errors). The last two types of errors show misunderstanding of period and possibly pupils just made a guess.

(ii) In Table 72, the first item is about the inverse of a function. The majority of pupils claimed wrongly the existence of such an inverse. The reason for this could be that they are familiar with the method of finding the inverse and the common error "$\sqrt{x^2} = x$" may lead them to such error.

(iii) The first question of Table 75 is about the number of roots of a quintic. The formulation of this item may be one factor in its difficulty, while misunderstanding of the discriminant is another.

(iv) The last item in Table 77 probably has many factors which affect pupils in reaching the right answer. This item tests, not just the subject matter, but also understanding of the formulation and diagram. Confusion may arise from many sources.

DI \ FV	below 0.40	[0.40 - 0.60]	above 0.60
above 0.40	difficult*	acceptable*	easy*
[0.30 - 0.39)	difficult	improvable	easy
[0.20 - 0.29)	difficult	marginal	easy
below 0.20	rejected	rejected	rejected

Table 68. The classification of items' difficulty. (*) These items would normally be acceptable items for a test.

content	Frq
trigonometry	10
vectors	6
inverse of a function	3
area of a shape	3
transformation	2
roots of a quadratic	2
gradient	2
integral	2
limit	2
circle properties	1
numeration	1
distance	1

Table 69. The classification of difficult items according to their content.

item (*)	the right answer and (its %)	main wrong answers and (their %)	possible causes of errors
1/13	sin 2x & 2 tan x (9)	2 sin x & 2 tan x (24) sin 2x & tan 2x (24) 2 sin x & tan 2x (21) other combinations (21)	similarity in form may lead to error; possible guess
2/36	$\surd(1 - a^2)$ (10)	$1 - a$ (62) $\frac{1}{a}$ (18)	the common errors $\sin(90 - \theta)^\circ = \sin 90^\circ - \sin \theta^\circ$ & $\surd(1 - a^2) = 1 - a$ lead to the first wrong answer; faulty recall of formulae
2/9	2π (10)	$\frac{\pi}{2}$ (44) $\frac{3\pi}{2}$ (26) π (14)	misunderstanding of period, using $\frac{\pi}{2} + \frac{3\pi}{2} = 2\pi$, and guessing are some reasons for wrong answers
2/32	$\theta = \frac{3\pi}{8}$ (34)	$\theta = \frac{3\pi}{4}$ (22) $\theta = \frac{\pi}{8}$ (16) $\theta = \frac{\pi}{2}$ (16)	finding the value of 2θ, misunderstanding of maximum

2/29	$\theta = \frac{\pi}{4}$ (38)	$\theta = \frac{3\pi}{4}$ (20) $\theta = \frac{5\pi}{4}$ (16) $\theta = \frac{\pi}{2}$ (16)	finding the value of 3θ, confusion between maximum and minimum, and faulty recall of values of cosine could be the reasons for these wrong answers
3/34	$\dfrac{\frac{2x}{y}}{1 - \frac{x^2}{y^2}}$ (37)	$\frac{x}{y} + \frac{w}{z}$ (32) $\dfrac{\frac{2x}{y}}{1 + \frac{x^2}{y^2}}$ (16)	the common errors $\tan(\theta_1 + \theta_2) = \tan\theta_1 + \tan\theta_2$ $\tan\theta = \frac{2\tan\frac{1}{2}\theta}{1 + \tan^2\frac{1}{2}\theta}$ may be the reasons for these worng answers
3/14	$-\sin(p - q)$ (34)	$\sin(p - q)$ (32) $-\sin(p + q)$ (28)	maybe carelessness and faulty recall of the $\sin(x + y)$ formula cause these wrong answers
4/27	$\tan^2 x = 1 - \frac{2}{p}$ (23)	$\tan^2 x = \frac{1}{2 + p}$ (32) $\tan^2 x = p(p + 1)$ (18) $\tan^2 x = 2 + \frac{1}{p}$ (14)	faulty recall of $\tan 2x$ formula may affect the result

4/33	2 (25)	-2 (33) -1 (25) $-\frac{1}{2}$ (17)	faulty recall of of the $\tan(x - y)$ formulae, guessing
4/30	-7 (39)	$-\frac{1}{7}$ (22) none of these values (22)	carelessness in calculation, choice of wrong quadrant

Table 70. The analysis of trigonometric items. (*) The item's notation gives the school then the item number.

item (*)	the right answer and (its %)	main wrong answers and (their %)	possible causes of errors
1/27	$\vec{BF} = -2\underline{p} + \underline{q}$ (23)	$\vec{BF} = -\underline{p} + \underline{q}$ (38) $\vec{BF} = -\underline{p} - \underline{q}$ (27)	$\vec{AF} = \pm\underline{q}$, $\vec{CF} = -\underline{p}$ failure to recall the $\vec{AB} = \vec{AC} + \vec{CB}$ or use the regular hexagon property
1/40	some other combination of responses (27)	(3) only (38) (2) only (21) (1), (2) and (3) (15)	failure to examine (3), faulty recall of collinearity condition

3/32	$\|\underline{a}\| = 9$ (36)	$\|\underline{a}\| = 5 + 2\sqrt{2} - 4\sqrt{3}$ (34) $\|\underline{a}\| = 15$ (16)	it is clear that the common errors $\|\underline{a}\| = \alpha + \beta + \gamma$, $\|\underline{a}\| = \|\alpha^2 + \beta^2 + \gamma^2\|$ where $\underline{a} = \alpha\underline{i} + \beta\underline{j} + \gamma\underline{k}$ have been made
4/40	$\underline{p}\cdot\underline{q} = \frac{3}{4}$ (33)	$\underline{p}\cdot\underline{q} = \frac{\sqrt{3}}{2}$ (33) $\underline{p}\cdot\underline{q} = \frac{3}{2}$ (25)	confusion between scalar product and multiplication of numbers, careless-ness in calculation
4/36	$a = -b$ (33)	$a = \frac{1}{2}b$ (25) $a = -2b$ (25) $a = b$ (15)	carelessness in factorisation of an expression
4/24	$xy = 4$ (32)	$y = x$ (36) $x + y - 3z = 0$ (23)	confusion between parallelism and either equality or perpendicularity

Table 71. The analysis of vector items. (*) The item's notation gives to the school then the item number.

item (*)	the right answer and (its %)	main wrong answers and (their %)	possible causes of errors
1/16	no inverse exists (6)	$f^{-1}(x) = \sqrt{(x - 1)}$ (77) $f^{-1}(x) = \sqrt{x} - 1$ (12)	maybe the common errors $\sqrt{x^2} = x$ and $\sqrt{(x + y)} = \sqrt{x} + \sqrt{y}$ affect the results
3/38	diagram (4) (35)	diagrams (2) and (4) (23) other combinations (19) diagrams (1) and (3) (16)	do misunderstanding of inverse and the negative form of the item lead to these wrong answers
4/22	$[gof]^{-1}(x) = x - 1$ (39)	$[gof]^{-1}(x) = x + 1$ (22) $[gof]^{-1}(x) = \frac{1}{2}(2x - 1)$ (22) $[gof]^{-1}(x) = \frac{1}{2}(2x + 1)$ (13)	maybe pupils were confused between *gof* and $[gof]^{-1}$, and also between *gof* and *fog*

Table 72. The analysis of items on inverse of functions. (*) The item's notation gives the school then the item number.

item (*)	the right answer and (its %)	main wrong answers and (their %)	possible causes of errors
1/39	$\frac{1}{8}\pi a^2$ (15)	$\frac{1}{4}\pi a^2$ (32) $\frac{1}{2}\pi a^2$ (24) indeterminable (24)	carelessness, confusion between radius and diameter, and faintness of 3rd semi-circle in the diagram may lead to such wrong answers
2/17	$\pi y(y - x)$ (26)	none of these responses (36) $\pi y(y + 2x)$ (22)	confusion between radius and diameter, incorrect cancellation, errors of sign
3/40	100 (26)	200 (41) 40 (19)	failure to recall $\frac{\theta}{180}\pi r = l$, confusion between length of arc and area of sector, difficulties in interpreting the question

Table 73. The analysis of items on areas of shapes. (*): The item's notation gives to the school then the item number.

item (*)	the right answer and (its %)	main wrong answers and (their %)	possible causes of errors
1/33	$(x + 4)^2 + (y + 6)^2 = 36$ (32)	$(x + 4)^2 + (y + 6)^2 = 18$ (35) $(x + 4) + (y + 6)^2 = 9$ (18)	multiplying the square of the radius by 2 rather than 2^2, forgetting the affect of the dilatation on the radius
1/38	$x = \frac{\pi}{6}$ (15)	$x = 0$ (29) $x = \frac{\pi}{4}$ (29) $x = \frac{\pi}{3}$ (15)	the difficulty of this item may lead pupils to guess, does "mapped on to itself" lead pupils to choose the answer $x = 0$? confusion between sin and cos, and between amplitude and period

Table 74. The analysis of transformation items. (*) The item's notation gives the school then the item number.

item (*)	the right answer and (its %)	main wrong answers and (their %)	possible causes of errors
1/26	3 (21)	4 or more (27) 2 (21) 1 (18) 0 (15)	misuse of the discriminant, the common error $\sqrt{x^2} = x$, the formulation of the item
1/34	$x_2 = 3 + \sqrt{2}$ (34)	$x_2 = 3 + \sqrt{2}$ & $\frac{b}{a} = 6$ (27) none of (1), (2), (3)(24)	the sum of the roots "is" $\frac{b}{a}$, finding the requires knowedge not normally possessed by pupils at this stage
2/26	7 (30)	5 (26) 3 (18) 4 (14) 6 (12)	the difficulty of this item may be unfamiliarity (leading pupils to guess)

Table 75. The analysis of items on roots of a quadratic and numeration. (*) The item's notation gives the school then the item number.

item (*)	the right answer and (its %)	main wrong answers and (their %)	possible causes of errors
2/24	$y = 8x + 13$ (32)	$y = 5x + 10$ (22) $y = 3x + 8$ (22)	using y rather than y' to find gradient, reading a badly printed power 3 as 2
4/19	$p^2 = -4a$ (36)	$a = -\frac{1}{4}$ (27) $a = -4$ (23)	carelessness in obtaining the gradients from the equations
1/35	$f(x) = \cos 2x$ (15)	$f(x) = 2x$ (27) $f(x) = \sin 2x$ & $f(x) = \cos 2x$ (24) $f(x) = 2x$ & $f(x) = \cos 2x$ (21) $f(x) = \sin 2x$ (15)	guesses?
4/38	$\frac{1}{8}[5^4 - 3^4]$ (36)	$\frac{1}{8}(5^4)$ (46)	careless evaluation at lower limit

Table 76. The analysis of items on gradient and definite integrals.

(*): The item's notation gives the school then the item number.

item (*)	the right answer and (its %)	main wrong answers and (their %)	possible causes of errors
2/35	non-existent (32)	$-\frac{1}{x^2}$ (32) $\frac{1}{x^2}$ (28)	+ misread as −, interchanged terms when rearranging the expression with a common denominator
4/39	75 (39)	undefined (30) 0 (17)	incorrect expansion of $(5 + h)^3$
3/25	$2 + 2\sqrt{3}$ (31)	$1 + 2\sqrt{3}$ (36) $4 + \sqrt{3}$ (12) $2 + \sqrt{3}$ (10)	omission of one radius, confusion between radius and diameter

Table 77. The analysis of items on limits and distance. (*): The item's notation gives the school then the item number.

item (*)	the right answer and (its %)	main wrong answers and (their %)	possible causes of errors
4/26	the circles meet on an axis is false. (39)	C_2 lies completely in the first quadrant is false (22) the common tangents are parallel is false (17) the circles cut in two points is false (13)	*y*-axis excluded from the first quadrant, negative form of the item

Table 78. The analysis of items on properties of a circle. (*): The item's notation gives the school then the item number.

An Experiment on Paper I

The analysis of all items which caused difficulties in paper I revealed some factors which may have contributed - to some extent - to these difficulties. As an attempt to find out the affect of such factors, an experiment was carried out among 95 Higher Grade pupils of four different schools. The test was composed of modified versions of twenty items. Fourteen of the original items were difficult (i.e. their facilty values (FVs) in the preliminary examination were below 0.4) and the remaining six were average (i.e. their FVs were between 0.4 and 0.6).

Note that in this test, there were no control items since the goal of this was just to get some insight into how such factors could affect item difficulty.

Changes were made to the language and type of items. In particular, the changes were as follows:

(i) negative items changed to positive ones;

(ii) wording made more familiar;

(iii) multiple-completion items changed to multiple-choice ones.

The number of items in the categories was 3, 9 and 8 respectively. The items and their changes can be found in Appendix 10.

The results

Tables 79, 80, 81 (a & b) and 82 give the results of the test for each school (two classes in School 3 were tested). In these tables, the pupils' responses and the facility value for each item are given. Table 83 shows the frequency of success (for each school) in each item and the facility values of each item: the facility value of the original item (FV_0) and that of the modified one (FV).

A highly significant correlation ($r = 0.67$, $p < 0.01$) was obtained between the two variables (FV_0 and FV) and a significant difference ($t = -1.82$, $p < 0.1$) between the means of the two variables. Even though the items were, in general, still difficult (or average) after the changes had been made, this supports the view that negative form, familiarity and item type affect performance.

A comparison between FV_0 and the average FV for items in each of the three categories (i.e. negative form, familiarity and type) has been made. This can be found in Tables 84, 85 and 86 respectively. In the first category, two items out of three produced a significant positive change in facility value. In the second category, three items produced a significant positive change and three produced a positive change that was almost significant at the 10% level. In the third category, one item produced a significant positive change, three produced a positive change that was almost significant at the 10% level and one (item 3) produced a negative change.

To sum up, it seems to me that changes were, in general, improvements since, from twenty items which were modified, there was just one case in which there was a significant fall in performance. The explanation of this situation is given in the analysis of some examples below.

The analysis of some examples

(i) The formulation of item 6 (in the first category) contains two statements, the first was in a negative form, whereas the second was in a positive form. Some candidates may have worked with just the last one. The modified item was expressed in positive form. As we expected, there was better performance on the modified item.

(ii) Although there is a link between the mathematics content and a real life situation in the formulation of item 12 (in the second category), the modified item in which the real life content was removed and a diagram added was preferred by pupils since it tests just the subject matter.

(iii) The apparently simpler modified item with the multiple-completion style replaced by a multiple-choice one did not produce a better performance.

(iv) The reason why multiple-choice questions can be easier than multiple-completion ones is clear with item 18 (in the third category) since to obtain the solution to the modified item it is only necessary to check the first two choices, but the solution to the original requires the checking of all choices. There was a much better performance on the modified item (where ignorance of tan was not a handicap).

P \ I	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1		C	E	D	E	B	B	D	E	E	D	C	B	B	C	E	A	B	D	B
2	A	C	D	E	D	B	B	B	D	C	C	B	C	B	B	C	D	E	D	D
3	A	D	A	C	E	C	B	D	A	E	D	B	C	B	A	E	D	A	D	B
4	B	D	B	C	B	B	B	C	A	E	C	D	A	D	B	B	B	B	E	C
5	A	c	A	E	A	C	B	C		C	E	B	E	B	D	B	A	D	D	E
6	B	D	A	D	C	B	E	D	A	C	E	D	D	D	E		C	B	E	C
7	A	C	E	D	C	B	B	E	A	E	D	C	A	B	A	E	B	B	E	B
8	D	C	A	E	A	C	B	C	B	A	D	C	E	B	A	D	E	B	E	D
9	A	C	A	C	E	C	A	B	B	C	A	D	A	B	C	C	D	B	E	B
10	A	D	D	E	B	B	E	C	A	C	D	C	E	B	B	E	A	B	B	D
11	A	D	C	D	A	B	D	B	C	E	D	D	E	B	B	E	D	B	A	E
12	A	C	E	E	E	B	E	A	C	C	A	A	E	B	E	D	A	D	E	E
13	A	C	C	D	D	B	E	C	D	E	A	C	C	B	A	C	E	B	B	B
14	B	B	B	C	A	B	A	D	C	B	E	D	A	E	B	C	B	B	D	D
RA	A	C	A	D	B	B	B	D	A	C	D	C	A	E	C	A	B	B	D	B
FV*	64	57	36	36	14	71	50	29	36	43	43	36	29	7	14	0	21	71	33	36

Table 79. The pupils' responses and FV for each item (School 1).

Key: *I* item, *P* pupil, RA right answer, FV facility value.

*: percentage.

P \ I	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	A	C	E	D	A	B	B	A	A	C	D	C	A	B	B	C	E	B	E	B
2	A	C	A	E	E	B	B	D	A	C	A	D	D	D	C	C	B	B	C	B
3	A	C	B	C	E	C	B	D	C	C	D	E	A	B	C	C	B	C	D	B
4	A	E	A	E	B	B	B	D	E	C	A	C	C	B	B	A	E	B	D	A
5	A	C	E	E	C	B	B	D	A	C	E	C	E	B	B	C	D	B	D	C
6	A	C	A	E	A	B	B	D	E	C	A	D	C	B	B	A	C	B	D	E
7	A	C	A	D	B	B	B	A	C	C	C	A		B	A	E	B	A		B
8	A		A		B	B	B	B	D	C	D	C	A	B	B	C	D	A	B	A
9	A	C	B	E	B	B	E	D	A	C	C	D	B	B	C	C	E	B	C	A
10	A	C	E	A	B	C	B	D	A	E	C	D	E	D		B	C	B	C	B
11	A	C	E	A	C	B	B	D	E	E	C	B	D	B	C	E	D	B	A	B
12	A	D	E	E	E	B	A	D	D	C	C	D	A	B	A	E	B	B	D	A
13	A	E	E	E	A	B	B	D	A	E	A	C	A	B	B	C	D	E	D	C
14	A	C	B	D	E	D	B	D	A	D	E	A	C	B	C	C	D	C	D	C
15	A	C	C	E	D	B	A	D	D	C	D	D	B	E	B	D	C	A	D	A
16	A	E	D	E	A	B	B	A	C	C	D	D	E	B	C	A	D	C	B	D
17	A	D	E	A	C	B	B	D	B	C	C	D	E	B	B	B	C	B	D	E
18	A	C	B	C	B	B	E	B	C	C	E	B	D	B	D	A	D	B	E	D
19	A	B	D	E	A	B	B	D	E	E	D	A	D	E	B	E	E	B	D	C
20	C	A	B	E	B	A	C	D	E	D	C	D	A	E	A	C	B	A	D	C
21	B	C	B	E	C	B	B	C	A	C	C	D	D	B	C	B	A	C	E	A
22	A	D	B	E	B	E	D	B	E	A	D	E	D	B	C	C	A	D	E	B
23	E	C	B	D	E	D	C	E	A	E	C	E	B	D	C	E	B	D	A	E

24	A	D	A		E	E	D	B	A	C	C	B	E	B	A	B	D	E	D	E
25	A	C	A	A	E	E	E	A		C										
26	E	C	B	B	C	C	D	C	A	E	A	B	C	B	A	D	C	D	E	D
27	B	A	B	E	C	D	C	E	C	A	E	D	E	E	A	D	A	D	E	D
RA	A	C	A	D	B	B	B	D	A	C	D	C	A	E	C	A	B	B	D	B
FV*	81	59	26	15	30	63	59	56	41	63	26	19	22	15	33	15	22	44	44	26

Table 80. The pupils' responses and FV for each item (School 2).

Key: *I* item, *P* pupil, RA right answer, FV facility value.

*: percentage.

P \ I	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	A	c	A	D	E	B	B	D	A	C	B	A	D	B	C	E	B	B	D	B
2	A	B	A	D	E	B	C	D	A	C	D	C	A	D	C	E	A	B	D	B
3	A	C	A	D	B	C	B	E	A	C	D	D	A	B	B	B	B	B	E	B
4	A	C	D	D	B	B	B	B	A	C	A	C	A	B	E	C	B	B	E	B
5	A	C	D	E	B	B	B	D	A	C	E	C	D	B	A	E	E	B	D	B
6	B	C	A	D	D	B	B	D	E	C	D	C	D	B	B	E	B	B	D	D
7	B	C	E	E	A	B	B	D	A	C	C	A	A	B	C	D	B	A	D	B
8	A	C	E	E	B		B	D	A	C	A	C	A	B	B	E	C	B	D	E
9	A	C	B	D	A	B	B	D	A	C	A	D	A	B	A	A	A	A	D	C
10	E	D	C	D	B	E	D	D	A	C	B	C	A	B	C	A	A	B	D	D
11	A	D	B	D	B	B	B	C	A	C	A	C	D	B	B	E	B	D	B	B
12	A	C	B	E	C	B	D	A	C	C	D	B	A	B	C	A	C	B	A	B
13	C	A	A	A	B	B	B	C	B	C	A	B	D	E	C	A	B	A		B
14	A	C	B	E	B	D	B	A	E	A	C	C	A	B	A	D	D	C	D	B
15	E	B	E	D	B	D	B	B	E	C	D	C	E	B	C	B	A	B	E	C
16	B	A	A	E	B	C	B	D	E	D	C	C	D	B	B	E	A	D	D	B
17	A	A	E	E	E	C	B	C	A	C	A	C	B	B	A	C	D	B	E	B
18	A	D	E	D	E	C	B	B	E	C	A	D	A	B	D	C	C	B	A	C
19	A	C	E	E	E	E	C	C	C	C	E	C	D	B	C	B	D	D	B	A
RA	A	C	A	D	B	B	B	D	A	C	D	C	A	E	C	A	B	B	D	B
FV*	68	58	32	53	53	53	79	47	58	89	26	63	53	5	42	21	37	63	53	63

Table 81(a). The pupils' responses and FV for each item (School 3).

P \ I	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	B							D	C											
2	B	C		E		C	E	D						B						
3	C	C	B	D	A	B	D	E	E	B				B		D	A	A	C	C
4	D	E	E	E	D	E	B	D	E	B		C		B			A	B	D	B
5	A	C	E	E	A	B	D	C	A	C		D		B		B	B	A	B	D
6	D	C	A	E	B	C	B	C	A	C		B		D		B	C	B	E	E
7	A	B	E	E	A	B	A	D	A	E		C		B		B	D	B	D	C
8	A	C	A	D	B	C	E	D	C	C		D		B		C	E	C	D	E
9	C	C	E	D	B	B	B	D	A	A		B		D		B	C	B	E	E
10	C	A	E	D	B	A	B	D	A	C		D		B		C	D	E	D	B
11	B	C	D	E	B	B	B	D	C	A		C		B		D	B	B	D	C
12	A	C	E	E	A	B	C	D	A	C		C		B		B	E	B	D	C
13	B	C	C	D	B	A	D	D	C	C		C		D		E	A	B	D	B
14	B	C	E	D	B	B	B	D	C	A		E		B		D	B	B	D	B
RA	A	C	A	D	B	B	B	D	A	C	D	C	A	E	C	A	B	B	D	B
FV*	29	71	14	43	50	50	43	79	43	43		36		0		0	21	57	57	29

Table 81(b). The pupils' responses and FV for each item (School 3).

Key: *I* item, *P* pupil, RA right answer, FV facility value.

*: percentage.

P \ I	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	A	A	B	D	B	B	B	D	A	C	D	C	E	B	C	A	B	B	D	E
2	A	C	A	C	B	B	B	E	A	C	D	C	D	B	C	E	A	B	E	D
3	A	C	A	D	A	B	E	D	E	C	D	C	E	B	D	E	A	B	D	E
4	A	C	A	D	A	C	B	A	A	C	D	C	C	B	D	D	B	A	E	D
5	A	A	E	E	B	B	C	D	B	C	A	C	A	B	B	D	B	A	D	C
6	A	A	A	D	B	B	E	B	A	C	A	D	B	B	B	C	D	B	C	A
7	D			D	A	B	B	C	A	E	D	C		D	C					
8	A	C	E	E	B	B	C	E	E	C	D	D	E	D	B	E	A	A	E	B
9	B	C	A	E	B	B	C	B	B	C	D	E	D	C	D	B	E	A	C	D
10	B	C	E	A	B	A	E	C	E	B	D		A	A	C					
11	D	E	D	C	B	A	A	B	E	C	A	C		D	B					
12			B	B	A	E	D	B	A	A	E	C	D	B						
13	B	C	E	D	B	B	B	D	A	C	D	B	A	B	B	E	A	B	D	B
14	A		B	D	A	B	B	C	A	C	D	C	A	B	A	B	E	B	D	C
15	A	C	E	D	A	C	B	D	A	C	C	C	E	B	A	E	A	B	D	C
16	B	A	A	C	B	B	C	D	B	E	C	D	A	D	B	A	E	B	E	B
17	A	E	A	E	A	B	E	D	C	E	A	D	E	B	A	E	D	B	D	E
18	A	A	A	C	D	B	D	C	E	B	D	A	E	D	E	C	E	A	B	B
19	A	C	D	D	E	A	A	C	A	E	E	D	A	B	A	E	A	A	B	E
20	B	C	A	E	D	C	B	B	B	A	C	C	B	C	B	B	D	B	B	C
21	A	E	D	A	E	D	E	D	E	C	B	D	D	B	E	E	D	C	D	E
RA	A	C	A	D	B	B	B	D	A	C	D	C	A	E	C	A	B	B	D	B
FV*	62	48	43	43	48	62	38	38	48	62	52	52	29	0	19	9	14	48	38	19

Table 82

sch \ I	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Sch 1	9	8	5	5	2	10	7	4	5	6	6	5	4	1	2	0	3	10	5	5
Sch 2	22	16	7	4	8	17	16	15	11	17	7	5	6	4	9	4	6	12	12	7
Sch 3a	13	11	6	10	10	10	15	9	11	17	5	12	10	1	8	4	7	12	10	12
Sch 3b	4	10	2	6	7	7	6	11	6	6		5		0		0	3	8	8	4
Sch 4	13	10	9	9	10	13	8	8	10	13	11	11	6	0	4	2	3	10	8	4
Total†	61	55	29	34	37	57	52	47	43	59	29*	38	26*	6	23*	10	22	52	43	32
FV (%)	64	58	31	36	39	60	55	50	45	62	36	40	32	6	28	11	23	55	45	34
FV_0 (%)	60	50	47	21	40	46	39	35	26	44	39	26	34	6	15	15	15	9	35	29

Table 83. The number of successful attempts, FV and FV_0 for each item in the exprimental test.

†: the total is out of 95, *: the number is out of 81.

sch \ item	6	11	19
Sch 1	71	43	36
Sch 2	63	26	44
Sch 3a	53	26	53
Sch 3b	50		57
Sch 4	62	52	38
mean & SD of FVs	59.8 8.4	36.8 12.9	45.6 9.2
FV_0	46	39	35
difference	S (5%)	NS	S (10%)

Table 84. FV_0 and FVs of items with negative form changed to positive form.

sch \ item	2	4	5	7	9	12	14	16	17
Sch 1	57	36	14	50	36	36	7	0	21
Sch 2	59	15	30	59	41	19	15	15	22
Sch 3a	58	53	53	79	58	63	5	21	37
Sch 3b	71	43	50	43	43	36	0	0	21
Sch 4	48	43	48	38	48	52	0	9	14
mean & SD	58.6	38.0	39.0	53.8	45.2	41.2	5.4	9.0	23.0
of FVs	8.2	14.2	16.6	16.2	8.4	16.9	6.2	9.3	8.5
FV_0	50	21	40	39	26	26	6	15	15
diff	S*	S*	NS	NS	S**	NS	NS	NS	NS

Table 85. FV_0 and FVs of items with change of formulation.

*: at 0.1, **: at 0.005.

Key: diff difference.

sch \ item	1	3	8	10	13	15	18	20
Sch 1	64	36	29	43	29	14	71	36
Sch 2	81	26	56	63	22	33	44	26
Sch 3a	68	32	47	89	53	42	63	63
Sch 3b	29	14	79	43			57	29
Sch 4	62	43	38	62	29	19	48	19
mean & SD	60.8	30.2	49.8	60.0	33.3	27.0	56.6	34.6
of FVs	19.3	10.9	19.2	18.9	13.6	12.8	10.9	17.0
FV_0	60	47	35	44	34	15	9	29
diff	NS	S*	NS	NS	NS	NS	S**	NS

Table 86. FV_0 and FVs of items with multiple-completion style changed to multiple-choice style.

*: at 0.05, **: at 0.001.

Key: diff difference.

CHAPTER SEVEN

The study of Paper II

In this paper, the same method of analysis which was adopted for Paper I was applied. This concerned the distribution of the test scores, facility value (FV), discrimination index (DI), reliability coefficient (r), and attempted to identify the causes of candidates' difficulties. In addition, classification of questions according to their structures and estimation of the demand of items has been made. The results of this analysis are compared with those of Paper II of the SCE examination.

The analysis of Paper II

1. School 1

The test was composed of ten questions. Table 87 shows the classification of these questions according to their structure. In this classification, the construction, appearance and the link between parts have been taken into account. Some examples which are typical of the various categories in this classification can be found in Appendix 11. Note that six of these questions have independent parts, while the remaining four have dependent parts.

The distribution of the test scores is given in Table 88, and Figure 31 illustrates the results. As we found in the analysis of Paper I, this distribution is, in general, normal. The top and the bottom 27% of the sample scored from 65 to 97 (out of 100) and from 15 to 42 respectively.

The FVs of the items were calculated and an estimation of their demands was made. These can be found in Table 89. The average FVs of the items which have the same demand are given in Table 90 and displayed in Figure 32. It is clear that the performance of pupils, in general, decreases when the demand of items increases with some exceptions (such as on the item with $Z = 6$ in which familiarity may explain the higher performance on it).

For each question and each part, the FV, DI, r and the evaluation of the items according to Macintosh's criterion are given in Tables 91 and 92 respectively. Once again, a strong correlation ($p < 0.01$) between the two FVs (of the original and new samples) was found. The evaluation of the items leads to a classification of them according to their difficulties. This can be found in Tables 93 and 94. It was found that, of the 10 questions, 9 were difficult, but two of them were acceptable. This may explain the very low value of r. Of the 25 parts, 15 were difficult but 6 of them were acceptable. The value of the reliability coefficient is high.

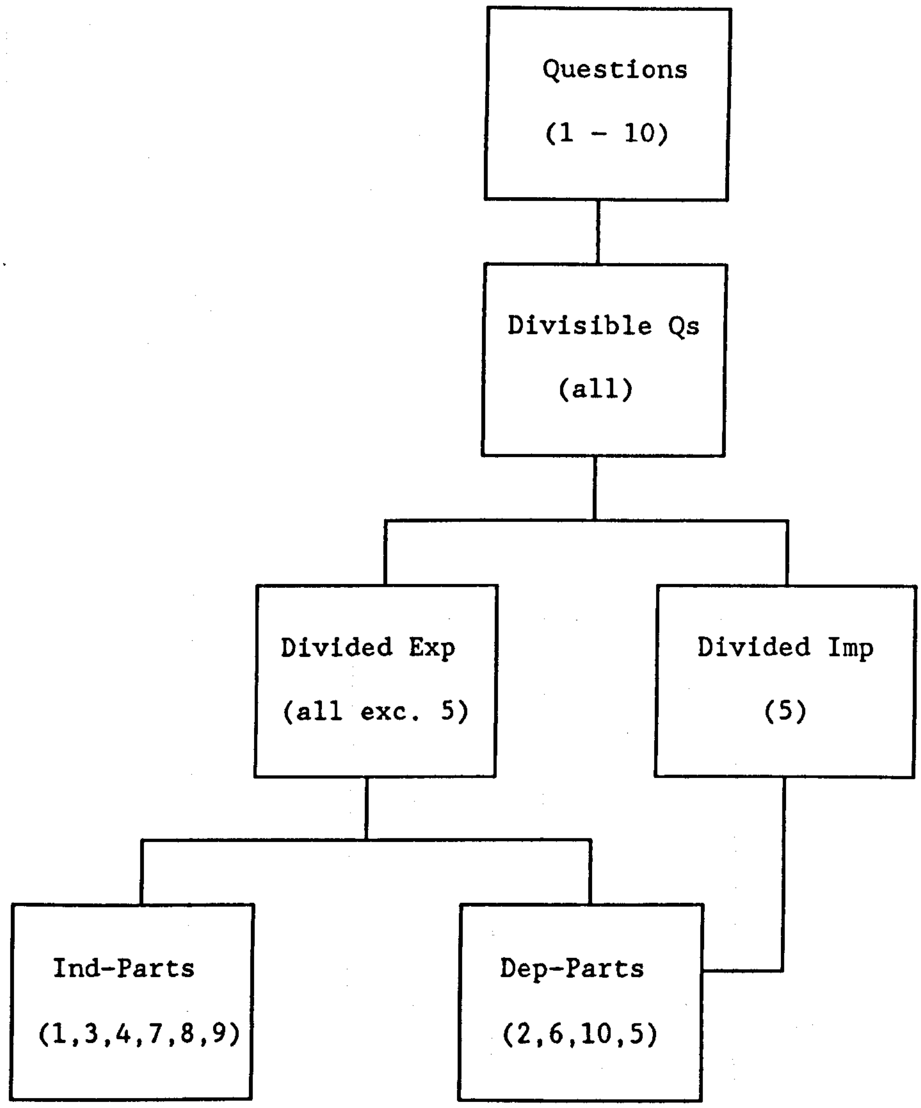

Table 87. The parts' division (School 1).

Key: Exp explicitly, Imp implicitly, Ind independent, Dep dependent.

Frq	Mark
1	97
1	87
1	80
1	72
1	70
1	67
2	66
1	65
1	64
1	62
1	60
1	57
3	56
1	55
1	51
2	50
1	48
3	47
1	43

← the top 27% of the sample ($N = 9$) [marks 97 to 65]

1	42
2	40
1	33
1	32
1	31
1	30
1	20
1	15

← the bottom 27% of the sample ($N = 9$) [marks 42 to 15]

Table 88. The distribution of test scores of School 1.

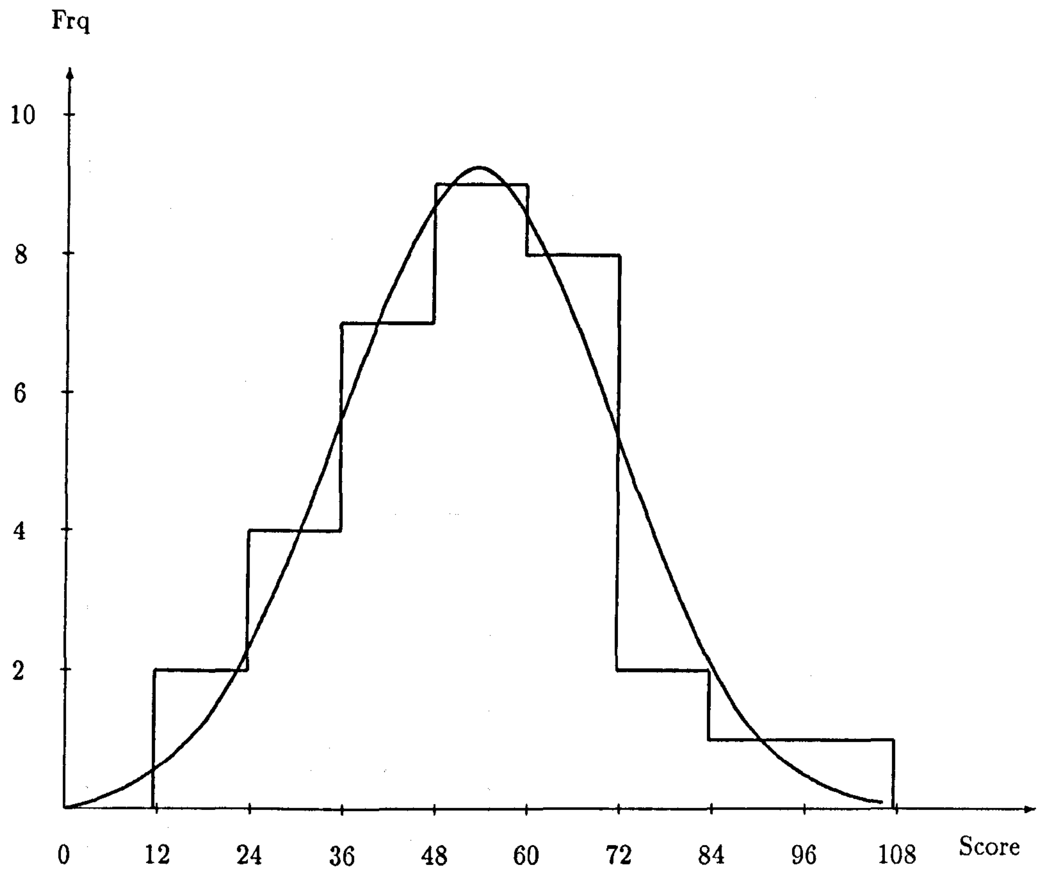

Figure 31

range 82

median 53

mean 53

SD 17.6

item		Frq	FV (%)	Z
2		11	32	11
3	a	20	59	6
	b	8	24	4
	c	2	6	5
4	a	12	35	4
	b	11	32	5
	c	7	21	5
5		16	47	9
6		4	12	12
7	a	9	27	5
	b	11	32	7
8	a	5	15	9
	b	24	71	5
	c	3	9	5
9	a	17	50	3
	b	1	3	7
10		7	21	11

Table 89/Sch 1

Frq	1	2	6	1	2	2	2	1
FV (%)	50	30	28	59	15	31	27	12
Z	3	4	5	6	7	9	11	12

Table 90. The average FVs of items which have the same demand.

item	FV_1	FV_2	DI	evaluation
1	0	0	0	rejected
2	0.32	0.33	0.44	difficult
3	0.03	0.06	0.11	rejected
4	0.09	0.17	0.33	difficult
5	0.47	0.50	0.55	acceptable
6	0.12	0.17	0.33	difficult
7	0.21	0.33	0.67	difficult
8	0.03	0.06	0.11	rejected
9	0.03	0.06	0.11	rejected
10	0.21	0.22	0.22	difficult

Table 91. The FV and DI for each question of School 1 ($r = 0.07$).

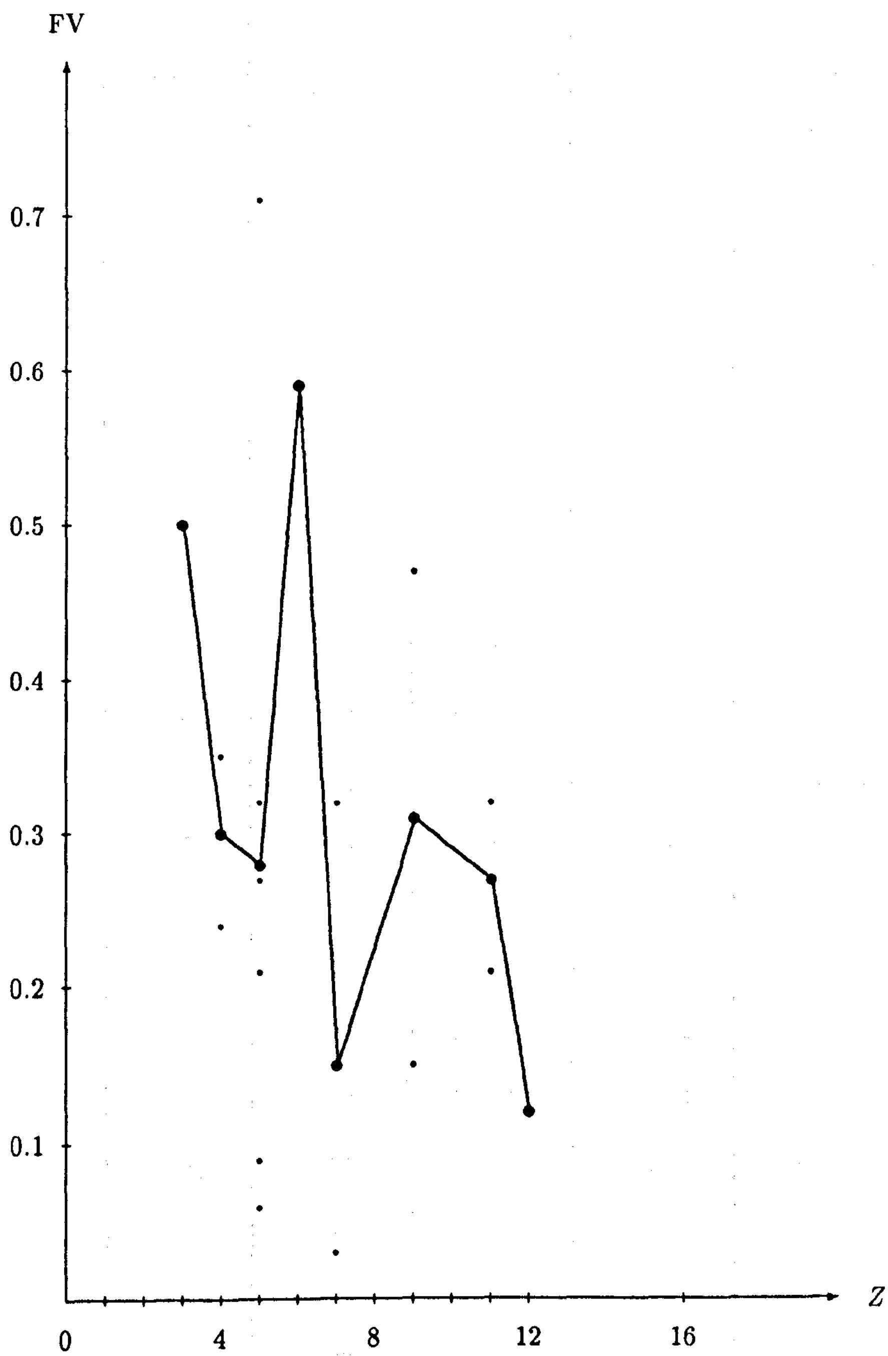

Figure 32

item	FV_1	FV_2	DI	evaluation
1a	0.03	0.06	0.11	rejected
1b	0.00	0.00	0.00	rejected
2a	0.56	0.50	0.55	acceptable
2b	0.50	0.56	0.67	acceptable
2c	0.47	0.44	0.44	acceptable
3a	0.59	0.56	0.67	acceptable
3b	0.24	0.17	0.33	difficult
3c	0.06	0.11	0.22	difficult
4a	0.35	0.39	0.33	difficult
4b	0.32	0.28	0.55	difficult
4c	0.21	0.22	0.44	difficult
5_1	0.85	0.83	0.11	rejected
5_2	0.50	0.50	0.55	acceptable
6a	0.85	0.78	0.44	easy
6b	0.71	0.67	0.44	easy
6c	0.12	0.17	0.33	difficult
7a	0.27	0.39	0.78	difficult
7b	0.32	0.39	0.78	difficult
8a	0.15	0.28	0.55	difficult
8b	0.71	0.67	0.67	easy
8c	0.09	0.11	0.22	difficult
9a	0.50	0.44	0.44	acceptable
9b	0.03	0.06	0.11	rejected
10a	0.21	0.22	0.22	difficult
10b	0.29	0.33	0.44	difficult

Table 92/Sch 1 ($r = 0.8$)

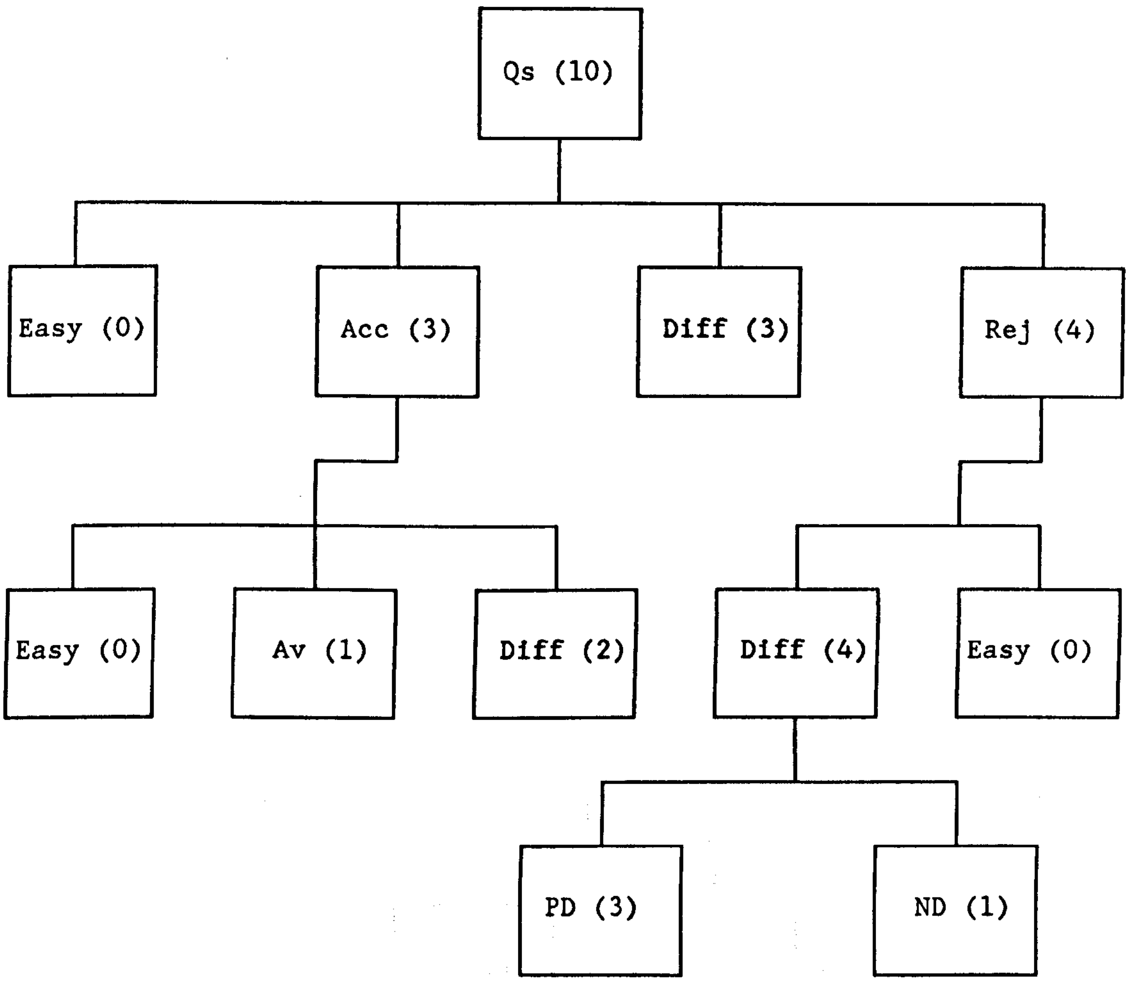

Table 93. The classification of questions' difficulty (School 1).

Key: Acc acceptable, Diff difficult, Rej reject, Av average, PD, ND poor, no discrimination, respectively.

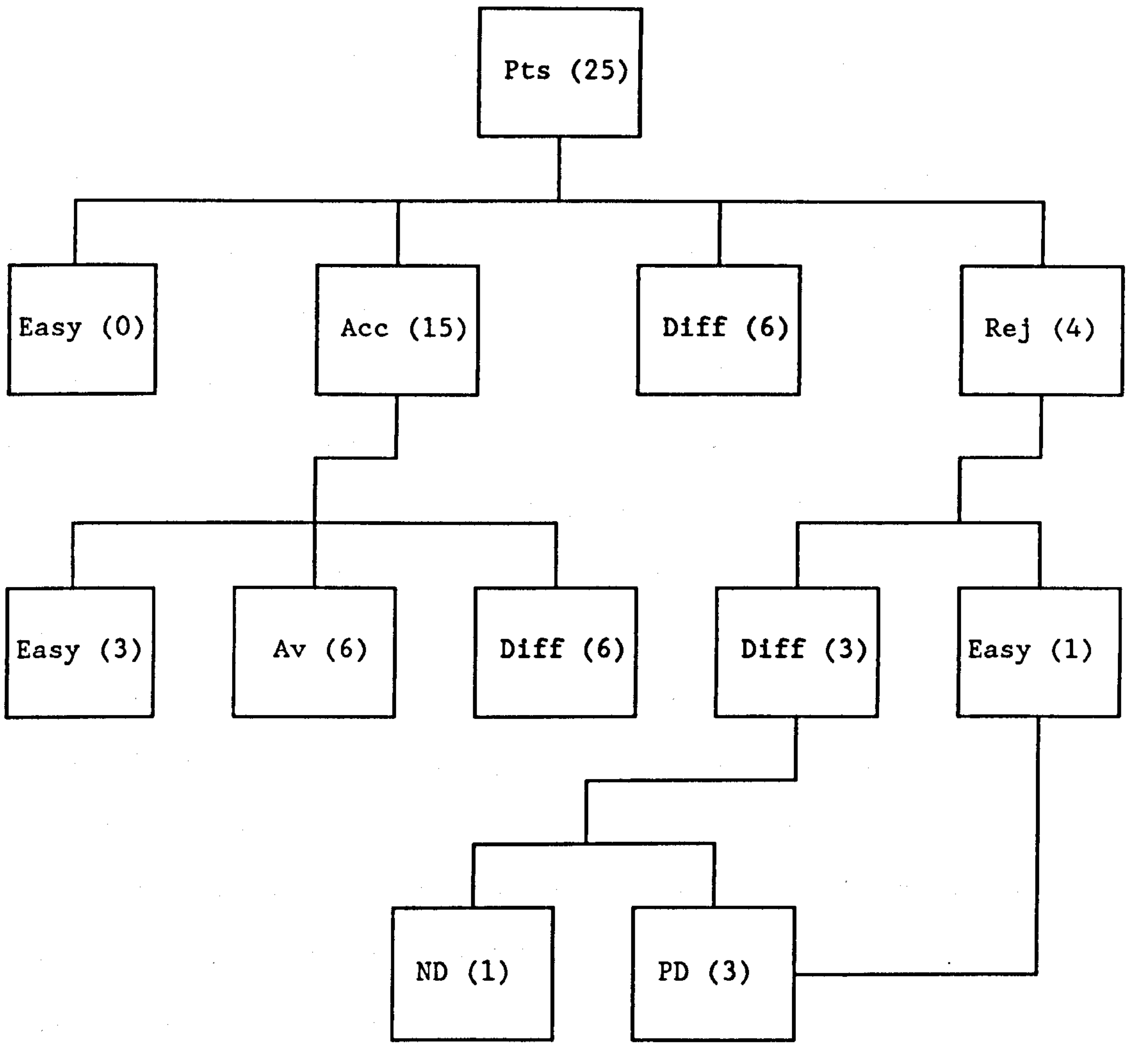

Table 94. The classification of parts' difficulty (School 1).

Key: Acc acceptable, Diff difficult, Rej reject, Av average, PD, ND poor, no discrimination, respectively.

2. School 2

The test was composed of 10 questions; their classification (according to their structure) is given in Table 95. The number of questions which have indivisible, independent and dependent parts was 2, 3 and 5 respectively (question 8 has a dependent part as well as independent ones).

The distribution of the test scores is given in Table 96 and illustrated in Figure 33. The top and the bottom 27% of the sample scored from 57 to 76 (out of 81) and from 5 to 34 respectively.

The FVs of items and an estimation of their demands are given in Table 97, whereas the average FVs of the items which have the same demand is shown in Table 98 and displayed in Figure 34.

Tables 99 and 100 give FV, DI, r and the evaluation of items as complete questions and as parts. A significant correlation ($p < 0.01$) between the two FVs was obtained. The evaluation of the items permits us to classify them according to their difficulties. This can be found in Tables 101 and 102. Note that 8 questions (out of 10) were found difficult but 4 of them were acceptable and the value of r was low. Of the 21 parts, 9 were difficult and 6 of them were acceptable. This may explain the high value of r.

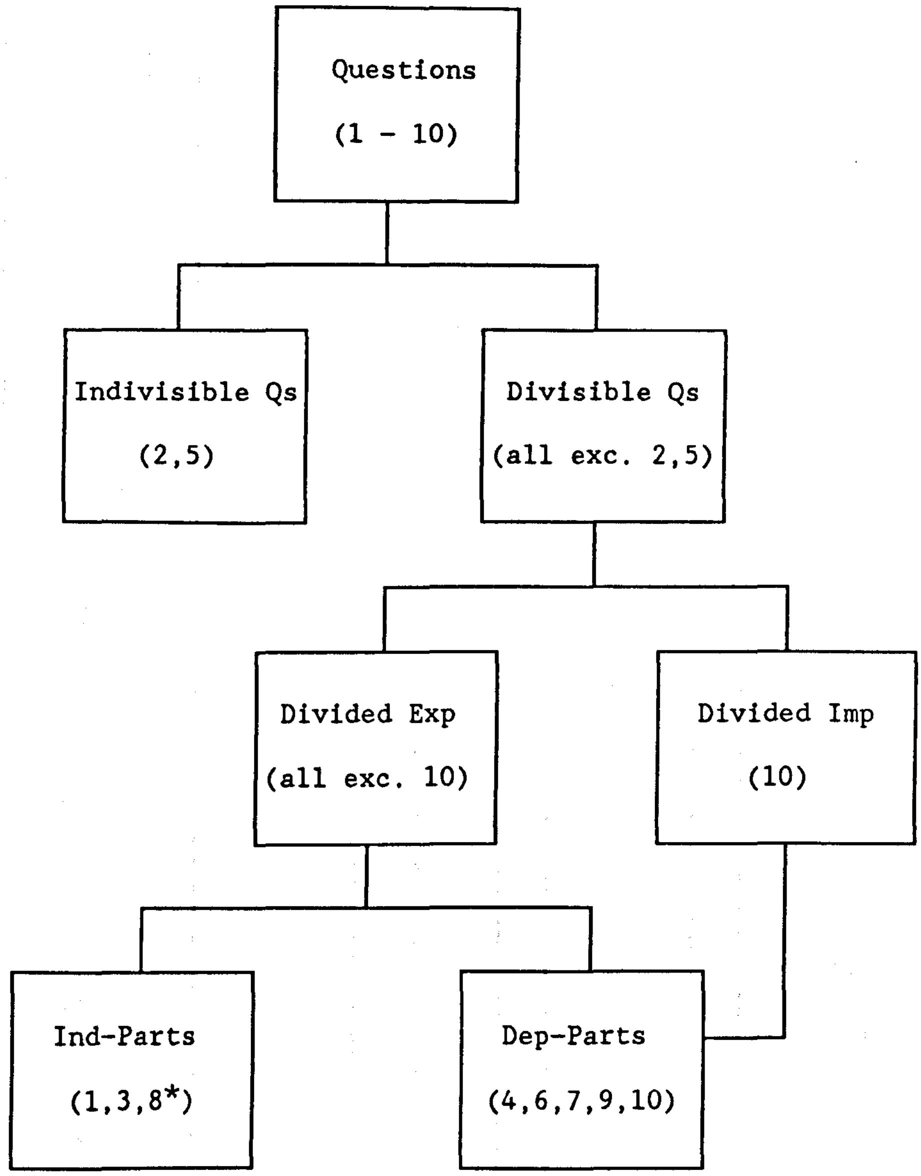

Table 95. The parts' division (School 2).

Key: Exp explicitly, Imp implicitly, Ind independent, Dep dependent. * item has a dependent part as well as independent ones.

Frq	Mark
2	76
2	72
1	69
1	67
1	63
2	62
2	61
3	57
1	56
1	53
2	52
2	48
1	46
2	45
1	44
1	43
3	42
2	41
1	40
1	39
2	38
1	37
1	35
2	34
1	33
1	29
2	28
1	26
1	25
1	24
1	22
1	21
1	18
1	6
1	5

Marks 76–57: ← the top 27% of the sample (*N* = 14)

Marks 34–5: ← the bottom 27% of the sample (*N* = 14)

Table 96. The distribution of the test scores of School 2.

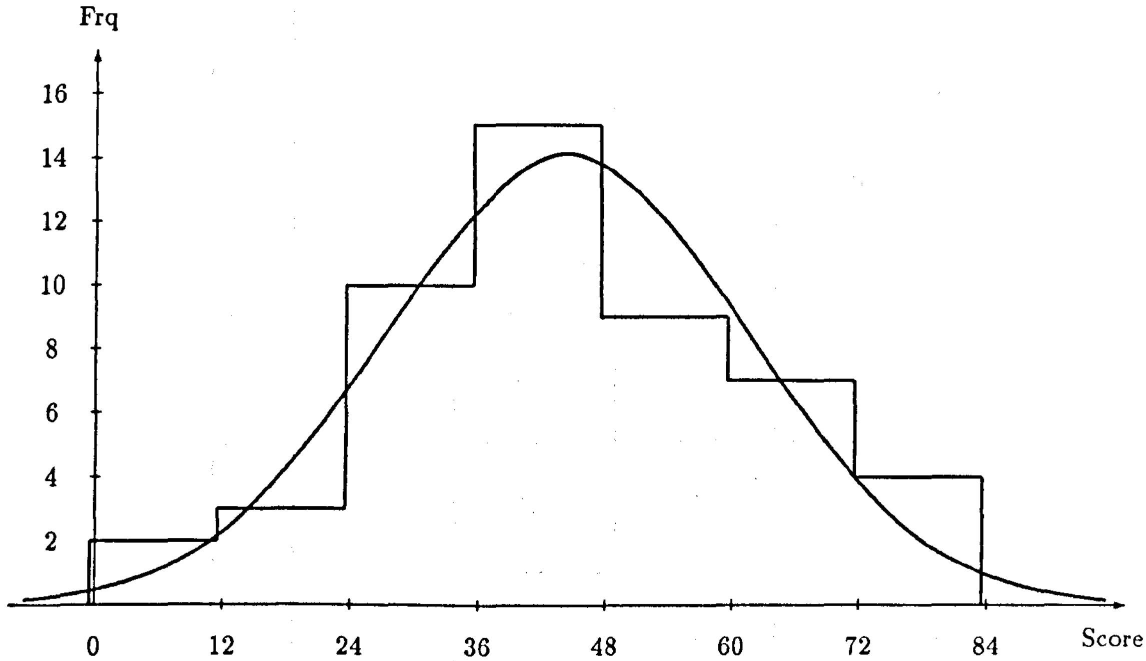

Figure 33

range 71

median 42.5

mean 44.2

SD 16.9

item		Frq	FV (%)	Z
2		18	36	6
3	a	21	42	5
	b	21	42	4
4		10	20	8
5		21	42	7
6		6	12	17
7		4	8	10
8	i	37	74	2
	ii	21	42	5
	iii	4	8	4
9		8	16	11
10		0	0	7

Table 97. The FV for each item of School 2.

Frq	1	2	2	1	2	1	1	1	1
FV (%)	74	25	42	36	21	20	8	16	12
z	2	4	5	6	7	8	10	11	17

Table 98. The average FVs of items which have the same demand.

item	FV_1	FV_2	DI	evaluation
1	0.42	0.43	0.57	acceptable
2	0.36	0.39	0.50	difficult
3	0.20	0.25	0.50	difficult
4	0.20	0.29	0.43	difficult
5	0.42	0.50	0.71	acceptable
6	0.12	0.21	0.43	difficult
7	0.08	0.14	0.29	difficult
8	0.06	0.11	0.21	difficult
9	0.16	0.25	0.36	difficult
10	0.00	0.00	0.00	rejected

Table 99. The FV and DI for each question of School 2 ($r = 0.40$).

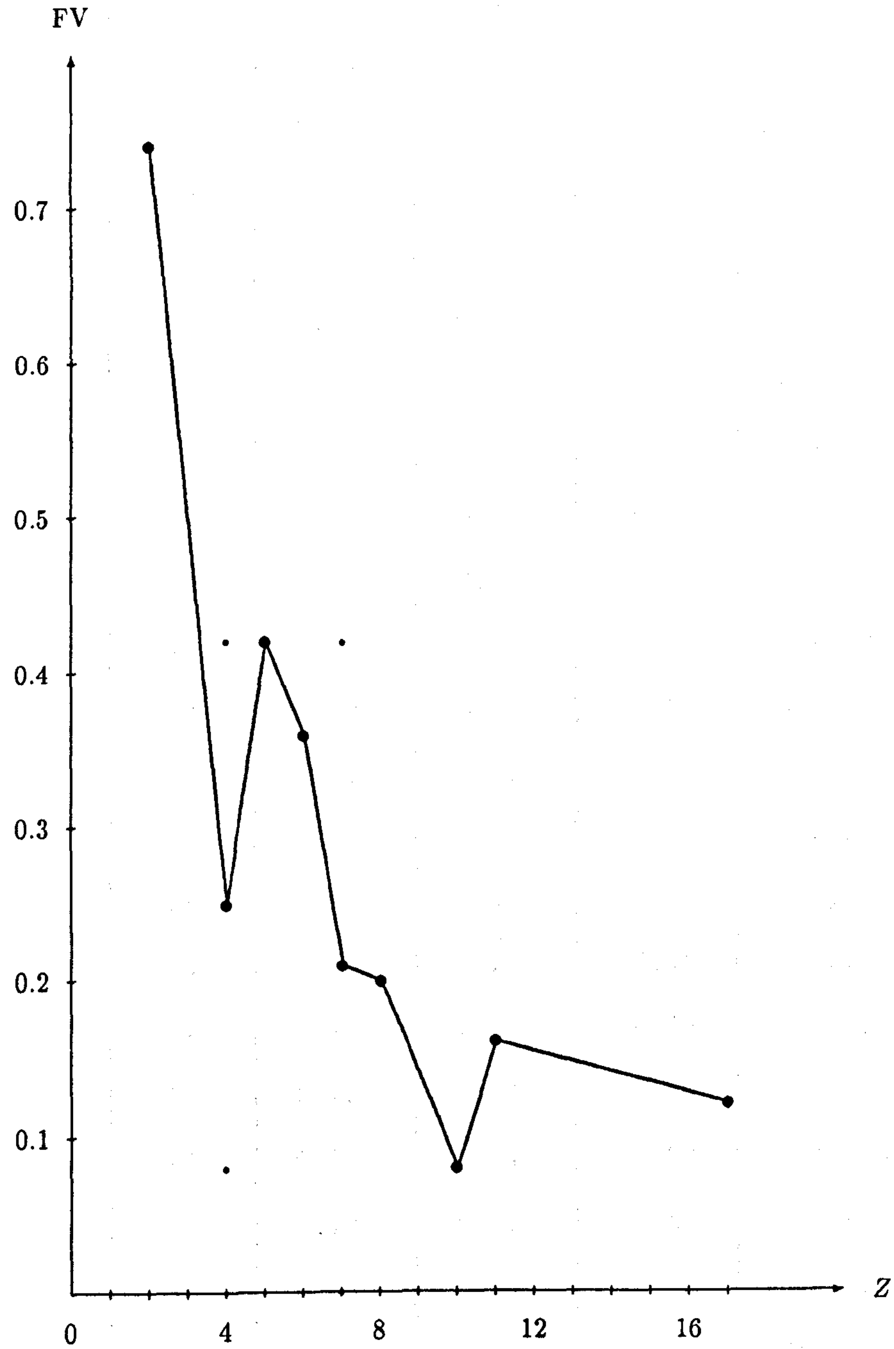

Figure 34

item	FV_1	FV_2	DI	evaluation
1a	0.56	0.57	0.43	acceptable
1b	0.76	0.68	0.50	easy
2	0.36	0.39	0.50	difficult
3a	0.42	0.32	0.50	difficult
3b	0.42	0.46	0.79	acceptable
4a	0.64	0.68	0.36	easy
4b	0.20	0.29	0.43	difficult
5	0.42	0.50	0.71	acceptable
6a	0.60	0.68	0.36	easy
6b	0.62	0.46	0.57	acceptable
6c	0.22	0.32	0.50	difficult
6d	0.22	0.32	0.64	difficult
7a	0.56	0.43	0.57	acceptable
7b/c	0.10	0.18	0.36	difficult
8i	0.74	0.68	0.36	easy
8ii	0.42	0.43	0.86	acceptable
8iii	0.08	0.14	0.29	difficult
$9a_1$	0.78	0.71	0.43	easy
$9a_2$	0.48	0.54	0.50	acceptable
9b	0.30	0.39	0.50	difficult
10	0.00	0.00	0.00	rejected

Table 100. The FV and DI for each item of School 2 ($r = 0.78$).

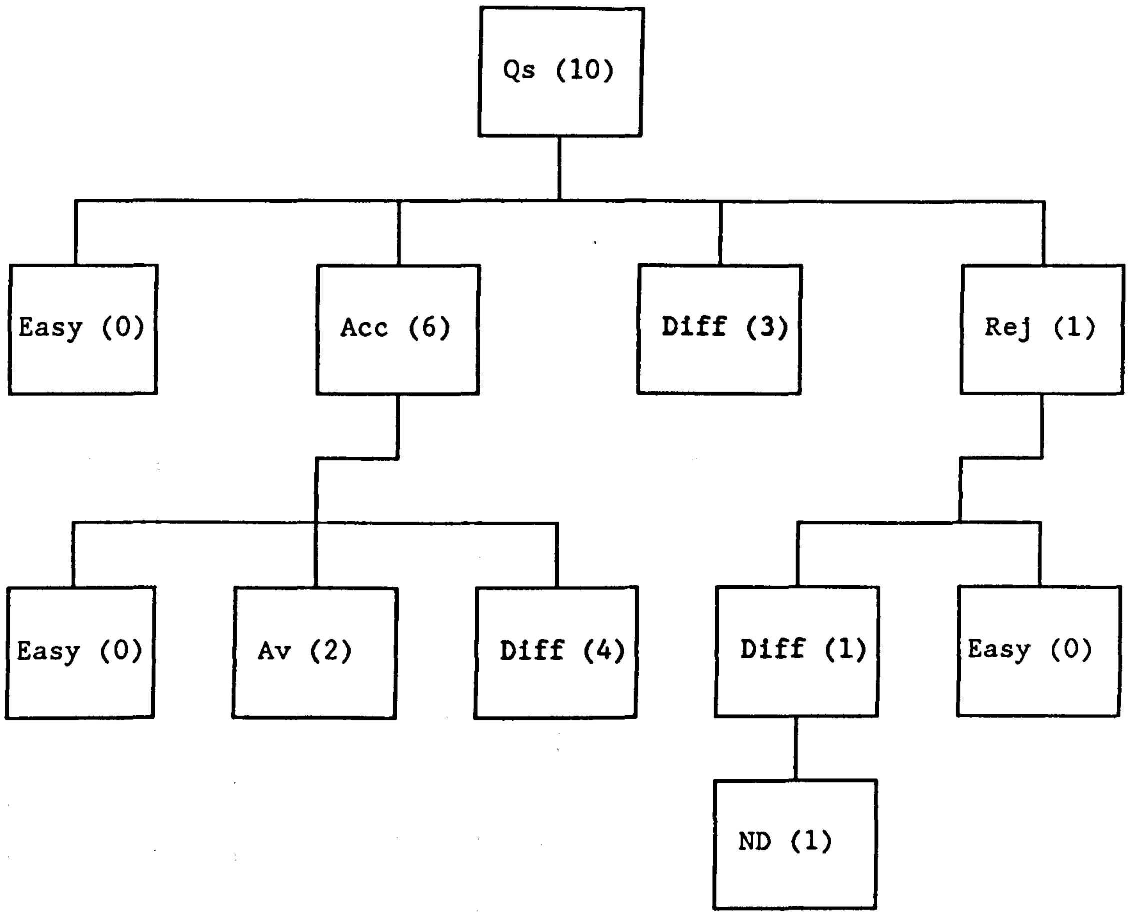

Table 101. The classification of questions' difficulty (School 2).
Key: Acc acceptable, Diff difficult, Rej reject, Av average, ND no discrimination.

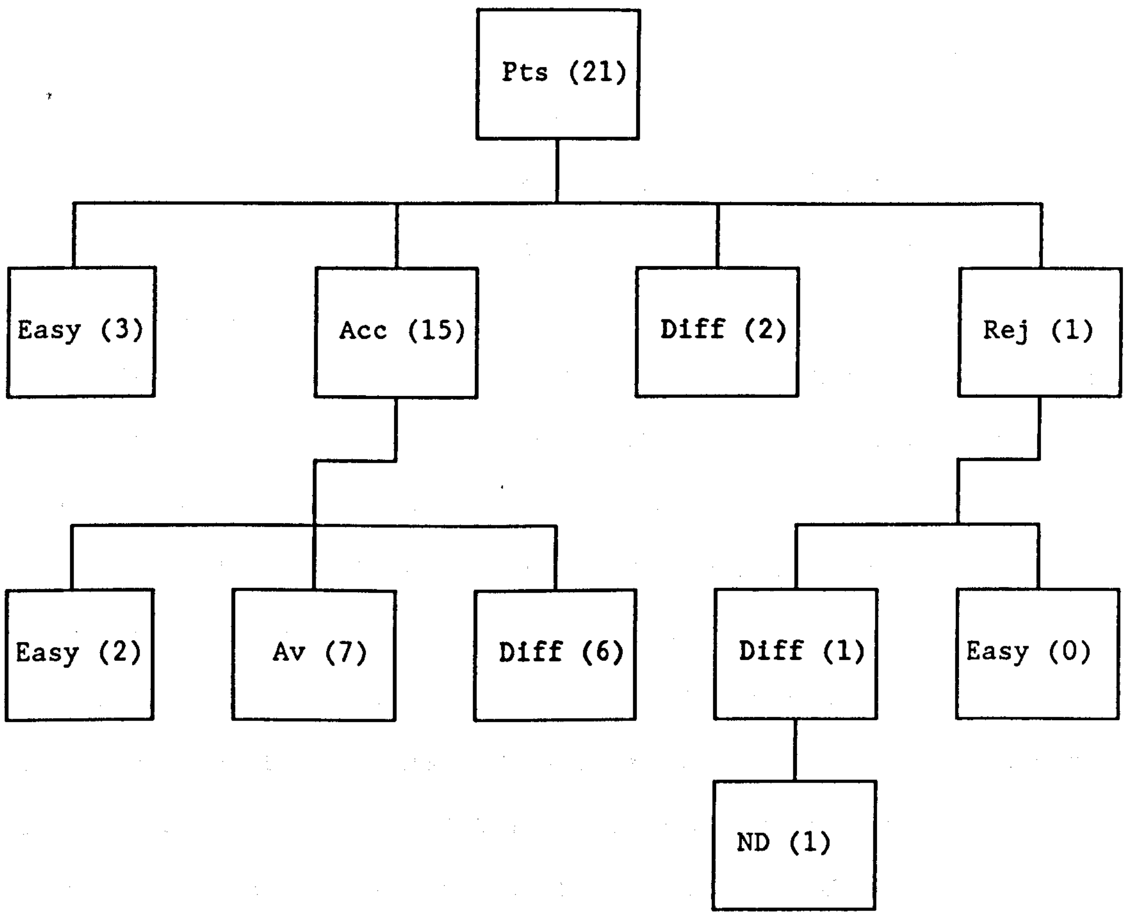

Table 102 The classification of parts' difficulty (School 2).

Key: Acc acceptable, Diff difficult, Rej reject, Av average, ND no discrimination.

3. School 3

The test was composed of 11 questions; their classification is given in Table 103. The number of questions which have indivisible, independent and dependent parts is 1, 7 and 3 respectively. (Question 10 is considered to have independent parts since we can solve the second part without solving the first.)

The test scores are given in Table 104 and illustrated in Figure 35. The top and the bottom 27% of the sample scored from 61 to 98 and from 6 to 32 respectively.

Table 105 shows the FVs of the items and their demands, while Table 106 gives the average FVs of the items which have the same demand. These are displayed in Figure 36.

The calculation of FV, DI, r and the evaluation of items in both questions and parts are given in Tables 107 and 108 respectively. A high correlation ($p < 0.01$) between the two FVs was obtained. Classification of questions and parts according to their difficulties is made in Tables 109 and 110 respectively. It was found that, all questions were difficult but 5 of them were acceptable. The value of r was relatively low. Out of 28 parts, 13 of them were difficult, of which 5 of them were acceptable. A high value of r was found.

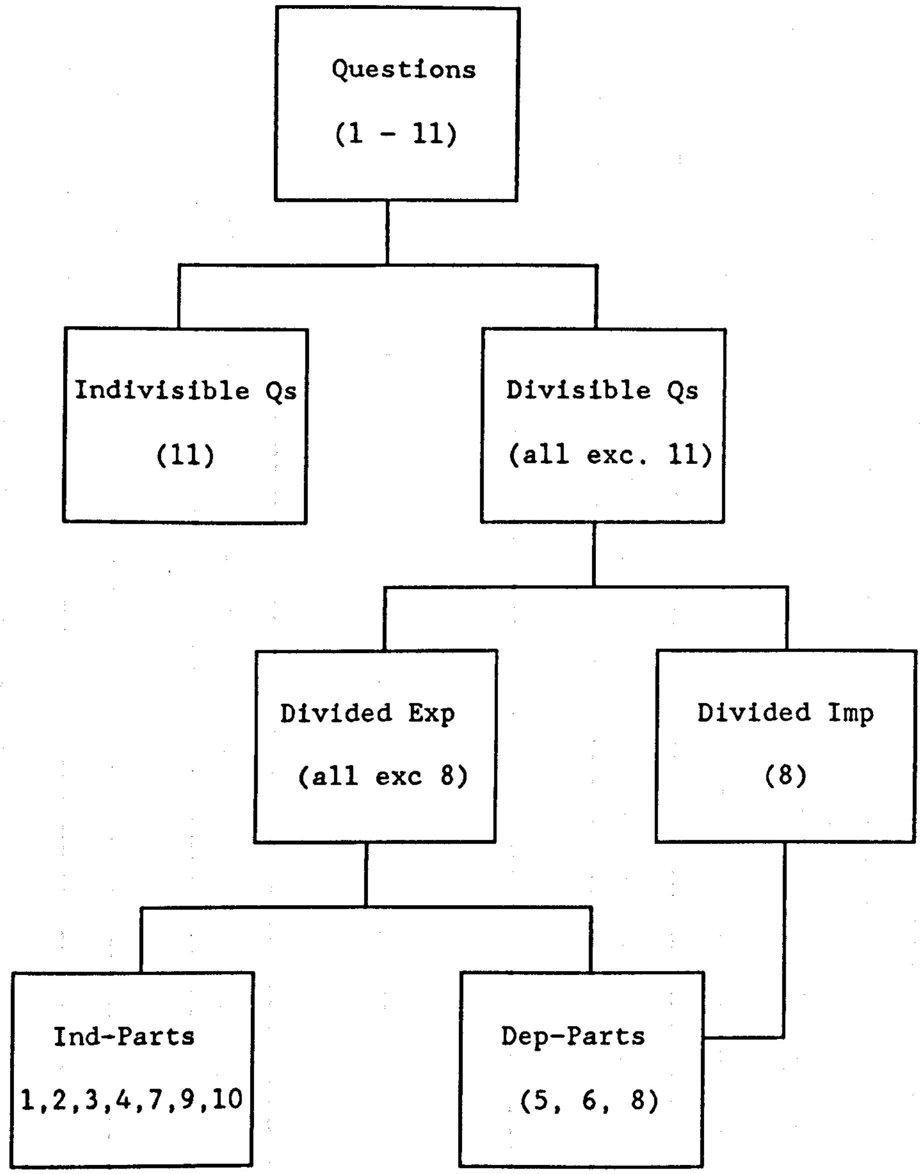

Table 103. The parts' division (School 3).

Key: Exp explicitly, Imp implicitly, Ind independent, Dep dependent. (Question 10 is considered to have independent parts.)

Frq	Mark
1	98
3	91
1	89
1	86
1	84
1	81
2	80
1	79
1	75
1	74
1	73
1	69
3	68
1	66
3	64
1	63
2	62
2	61
2	60
2	59
3	58
1	56
1	55
2	54
2	53
1	52
2	51
1	50
1	49
2	48
4	47
1	45
4	44
2	43
3	40
1	39
1	38
3	36
1	34
1	33
4	32
2	31
1	30
1	29
1	28
2	26
2	25
2	22
2	21
2	20
2	19
1	17
1	14
1	13
2	12
1	11
1	6

Table 104. The distribution of the test scores of School 3. The top and the bottom 27% of the the sample are indicated by ➡ (N = 26 for each of them).

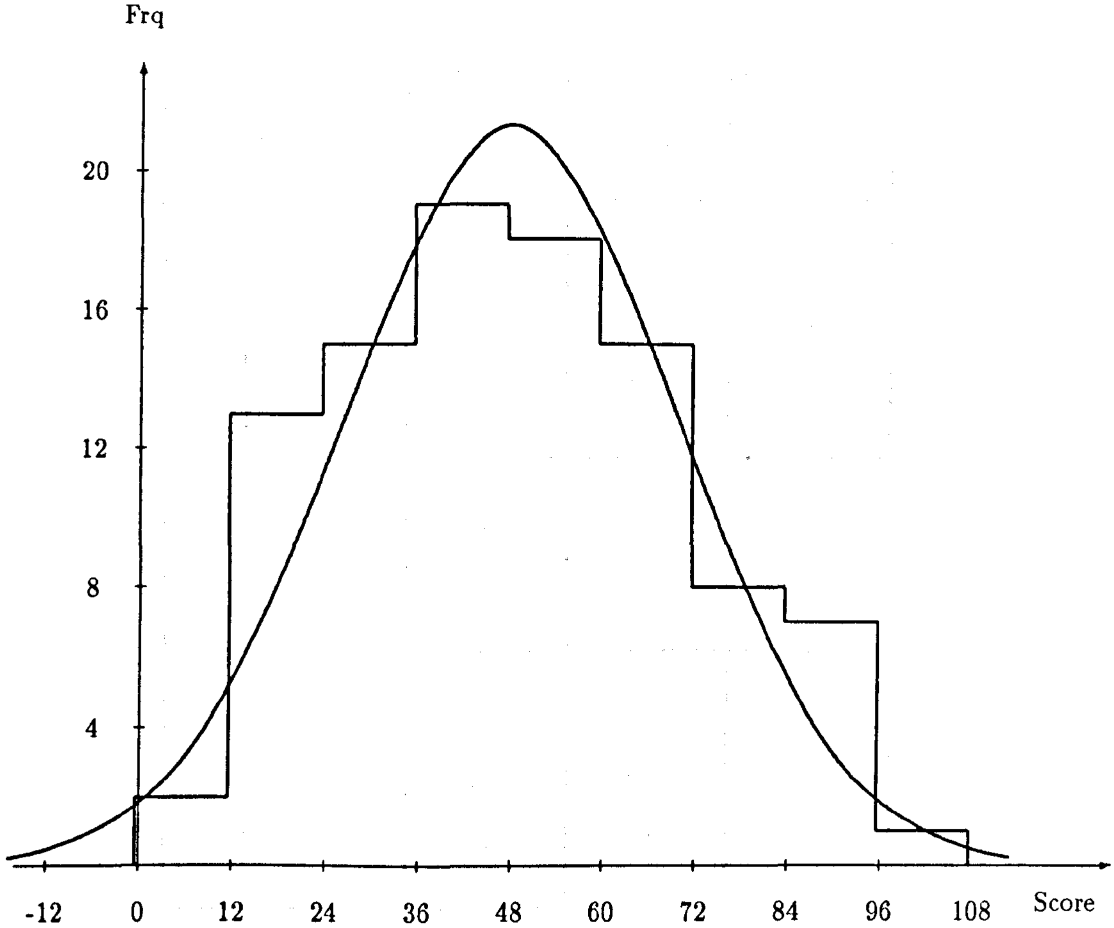

Figure 35

range 92

median 47

mean 47.8

SD 21.6

item		Frq	FV (%)	Z
2	a	50	52	6
	b_1	56	58	5
	b_2	32	33	2
3	a	51	53	6
	b	9	9	5
4	a	54	56	4
	b	12	13	7
	c	31	32	5
5		26	27	12
6		18	19	12
7	a_1	25	26	4
	a_2	39	41	2
	b	47	49	4
	c	29	30	5
8		27	28	9
9	a	5	5	9
	b	0	0	7

10	i ii	11 13	11 14	4 6
11*				

Table 105. The FV for each item of School 3.

* omitted because of an error in the item.

Frq	2	4	4	3	2	2	2
FV (%)	37	36	32	40	7	16	23
Z	2	4	5	6	7	9	12

Table 106. The average FVs of items which have the same demand.

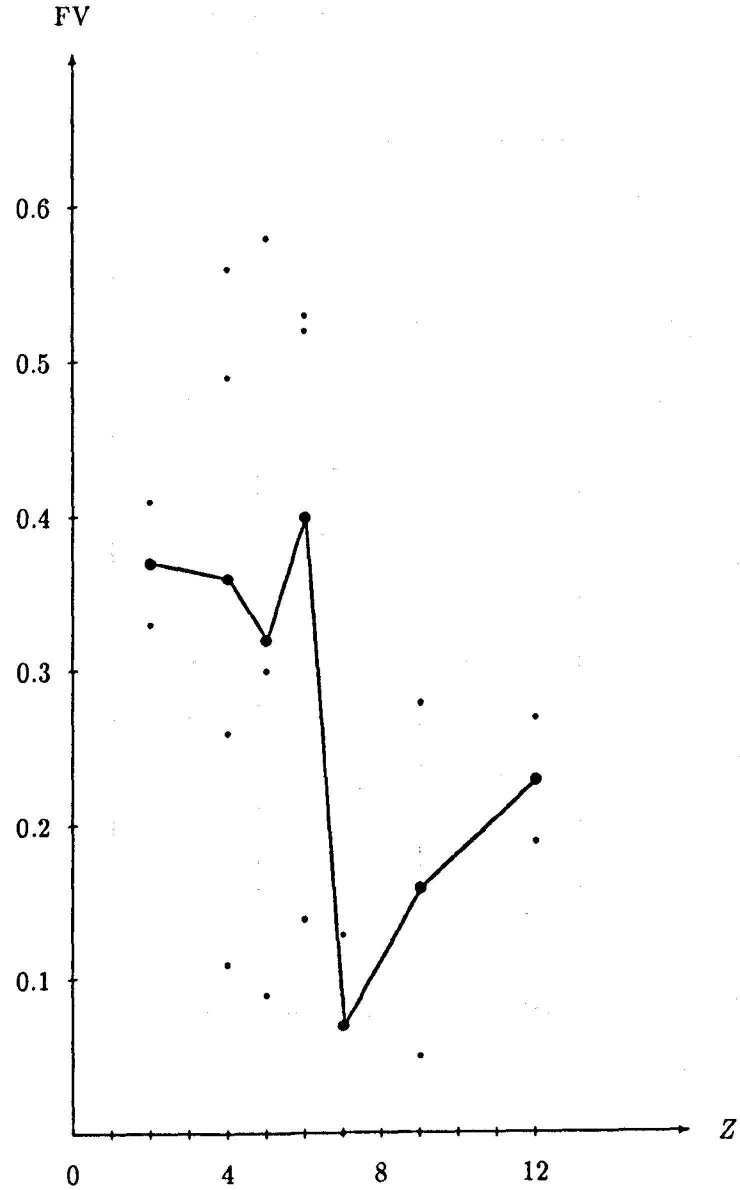

Figure 36

item	FV_1	FV_2	DI	evaluation
1	0.24	0.27	0.54	difficult
2	0.23	0.35	0.62	difficult
3	0.09	0.17	0.35	difficult
4	0.08	0.13	0.27	difficult
5	0.27	0.37	0.73	difficult
6	0.19	0.27	0.54	difficult
7	0.06	0.12	0.23	difficult
8	0.28	0.33	0.58	difficult
9	0.00	0.00	0.00	rejected
10	0.04	0.08	0.15	rejected
11	0.07	0.12	0.23	difficult

Table 107. The FV and DI for each question of School 3 ($r = 0.51$).

item	FV_1	FV_2	DI	evaluation
1a	0.39	0.37	0.65	difficult
1b	0.55	0.52	0.65	acceptable
1c	0.55	0.52	0.58	acceptable
2a	0.52	0.58	0.62	acceptable

$2b_1$	0.58	0.65	0.62	easy
$2b_2$	0.33	0.44	0.65	acceptable
3a	0.53	0.50	0.85	acceptable
3b	0.09	0.17	0.35	difficult
4a	0.56	0.52	0.65	acceptable
4b	0.13	0.19	0.38	difficult
4c	0.32	0.38	0.69	difficult
5i	0.38	0.42	0.85	acceptable
5ii	0.38	0.40	0.81	acceptable
6	0.19	0.27	0.54	difficult
$7a_1$	0.26	0.29	0.27	difficult
$7a_2$	0.41	0.38	0.38	difficult
7b	0.49	0.54	0.69	acceptable
7c	0.30	0.35	0.54	difficult
8_1	0.80	0.69	0.46	easy
8_2	0.60	0.48	0.50	acceptable
8_3	0.38	0.44	0.73	acceptable
$9a_1$	0.66	0.67	0.35	easy
$9a_{21}$	0.46	0.44	0.35	improvable
$9a_{22}$	0.06	0.12	0.23	difficult
9b	0.00	0.00	0.00	rejected
10i	0.11	0.21	0.35	difficult
10ii	0.14	0.21	0.42	difficult
11	0.07	0.12	0.23	difficult

Table 108. The FV and DI for each item of School 3 ($r = 0.86$).

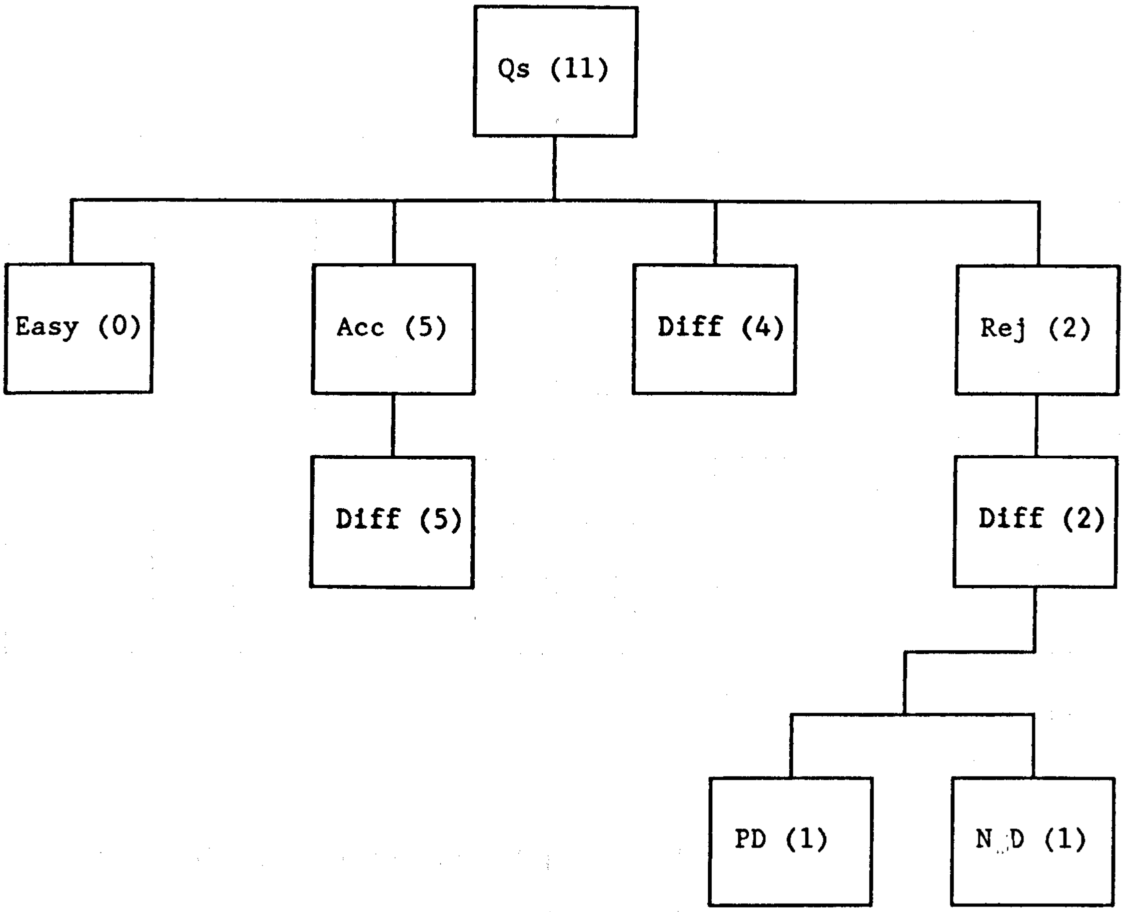

Table 109. The classification of questions' difficulty (School 3).

Key: Acc acceptable, Diff difficult, Rej reject, PD, NgD poor, no discrimination, respectively.

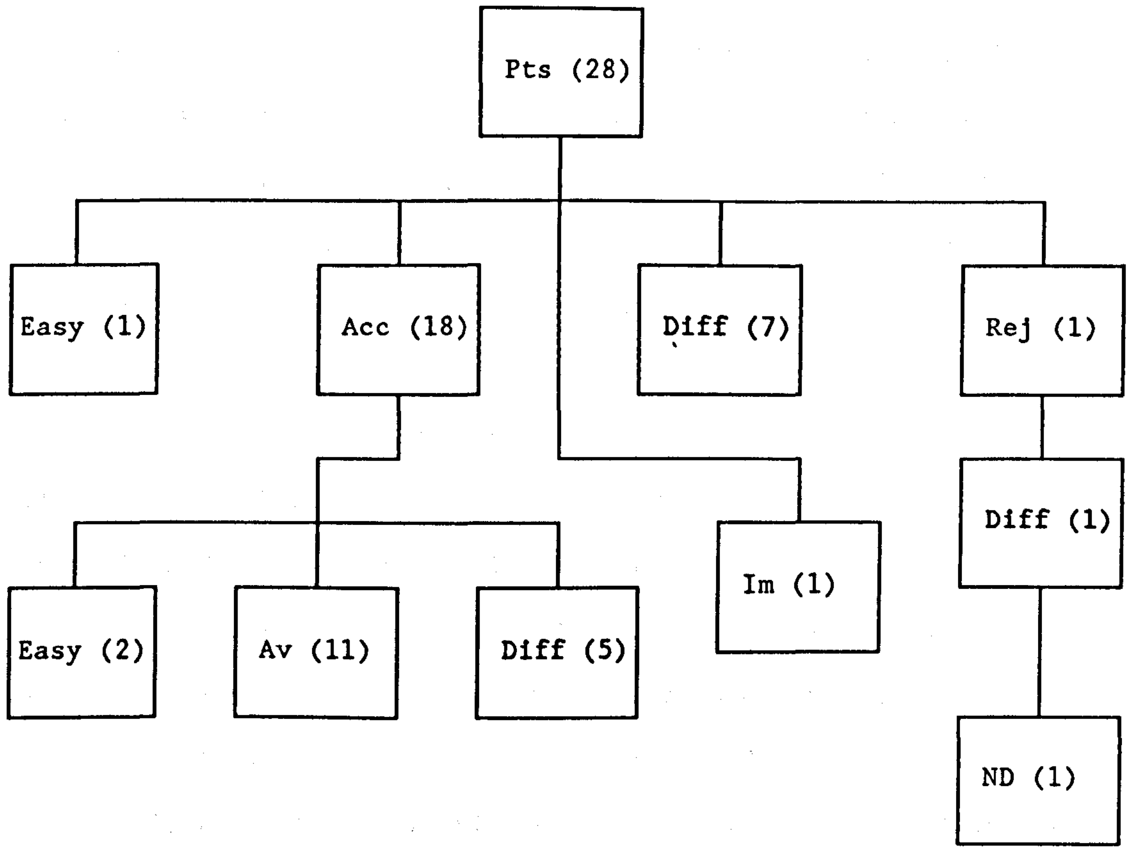

Table 110. The classification of parts' difficulty (School 3).

Key: Acc acceptable, Im improvable, Diff difficult, Rej reject, Av average, ND no discrimination.

4. School 4

The test was composed of 12 questions; 5 of them are not the same (at least in one part) for the two classes of the sample. Therefore it was decided to treat them separately (giving 17 items rather than 12). Classification of these items is given in Table 111. The number of the questions which have indivisible, independent and dependent parts was 1, 8 and 8 (item 11* has an independent parts as well as dependent one).

The distribution of the test scores is given in Table 112 and illustrated in Figure 37. The top and the bottom 27% of the sample scored from 74 to 85 and from 19 to 41 respectively.

The FVs of the items and their demands are shown in Table 113, while the average FVs of the items which have the same demand are given in Table 114 and illustrated in Figure 38.

FV, DI, *r* and the evaluation of items as complete questions and as parts are shown in Tables 115 and 116 respectively. Classification of the items according to their difficulties is given in Tables 117 and 118. Note that, of the 17 items, 14 were difficult but 5 of them were acceptable. Of the 34 parts, 14 were difficult but 3 of them were acceptable.

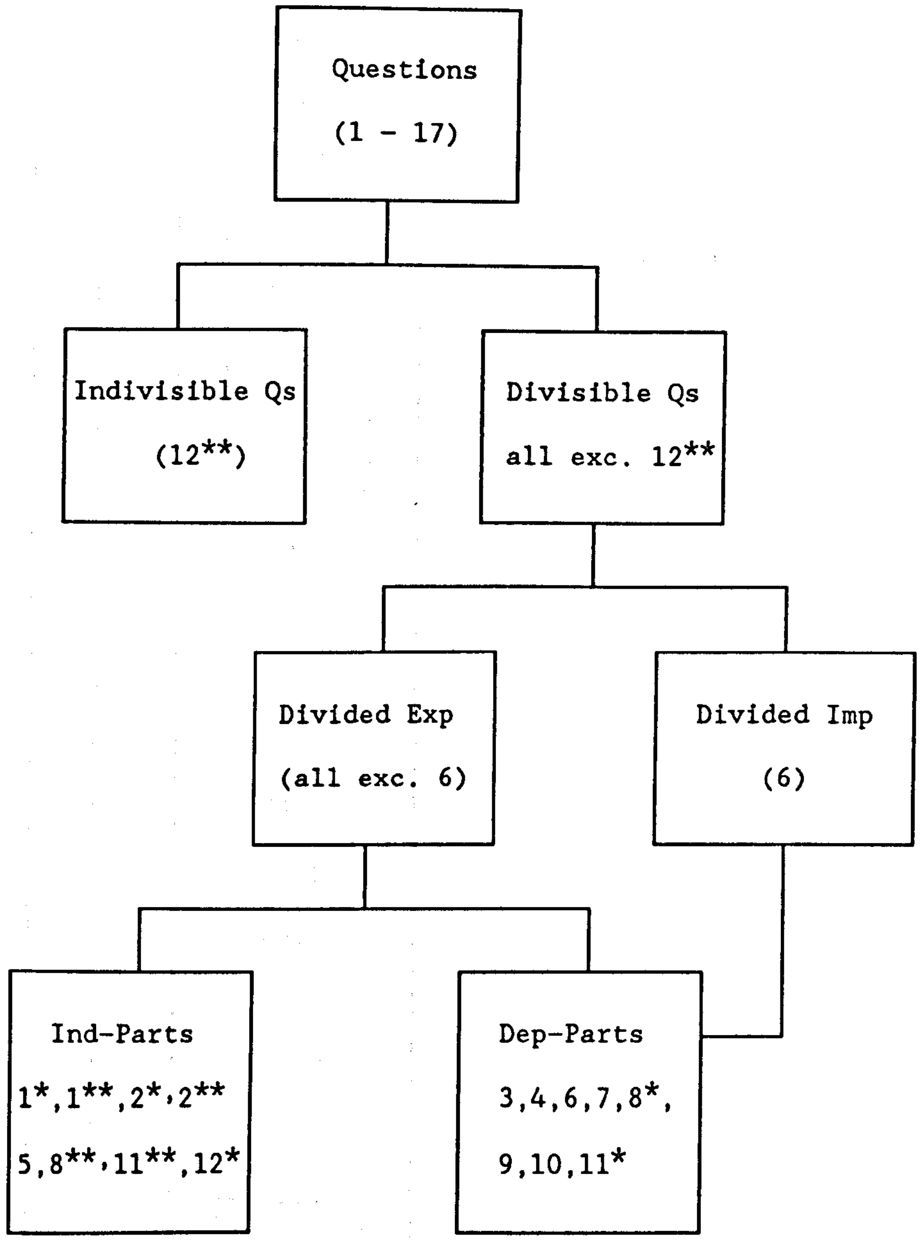

Table 111. The parts' division (School 4).

Key: Exp explicitly, Imp implicitly, Ind independent, Dep dependent. */** for the first/second class only.

Frq	Mark	
1	85.0	
1	83.0	
1	81.0	← the top 27% of the sample (*N* = 6)
1	80.0	
1	78.0	
1	74.0	
1	72.5	
1	72.0	
1	70.5	
1	70.0	
1	67.0	
1	66.5	
1	49.0	
1	47.0	
1	46.0	
1	44.0	
3	41.0	
1	37.0	
1	35.0	← the bottom 27% of the sample (*N* = 6)
1	28.0	
1	19.0	

Table 112. The distribution of the test scores of School 4.

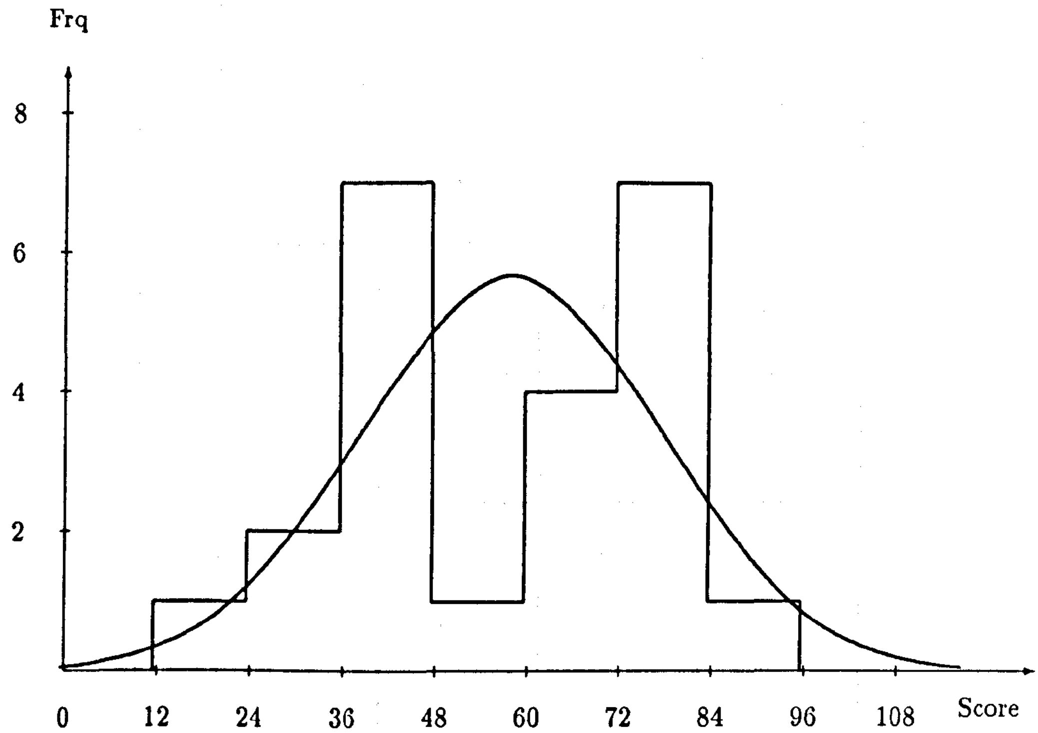

Figure 37

range 66

median 66.5

mean 57.7

SD 19.4

item		Frq	FV (%)	Z
2	a	0	0	5
	b*	4	33	5
	b**	1	9	4
3		4	17	11
4		6	26	8
5	a	8	35	6
	b	10	44	4
6		9	39	10
7		10	44	7
8	8*	2	17	12
	a**	4	36	3
	b**	0	0	8
9		2	9	10
10		1	4	13
11	11*/a**	11	48	4
	b**	9	82	3

	c_1**	2	18	3
	c_2**	6	55	5
12	a*	1	8	7
	b*	0	0	9
	T**	4	36	4

Table 113. The FV for each item of School 4.

*/** for the first/second class only.

Frq	3	4	3	1	2	2	1	2	1	1	1
FV (%)	45	34	29	35	26	13	0	24	17	17	4
Z	3	4	5	6	7	8	9	10	11	12	13

Table 114. The average FVs of items which have the same demand.

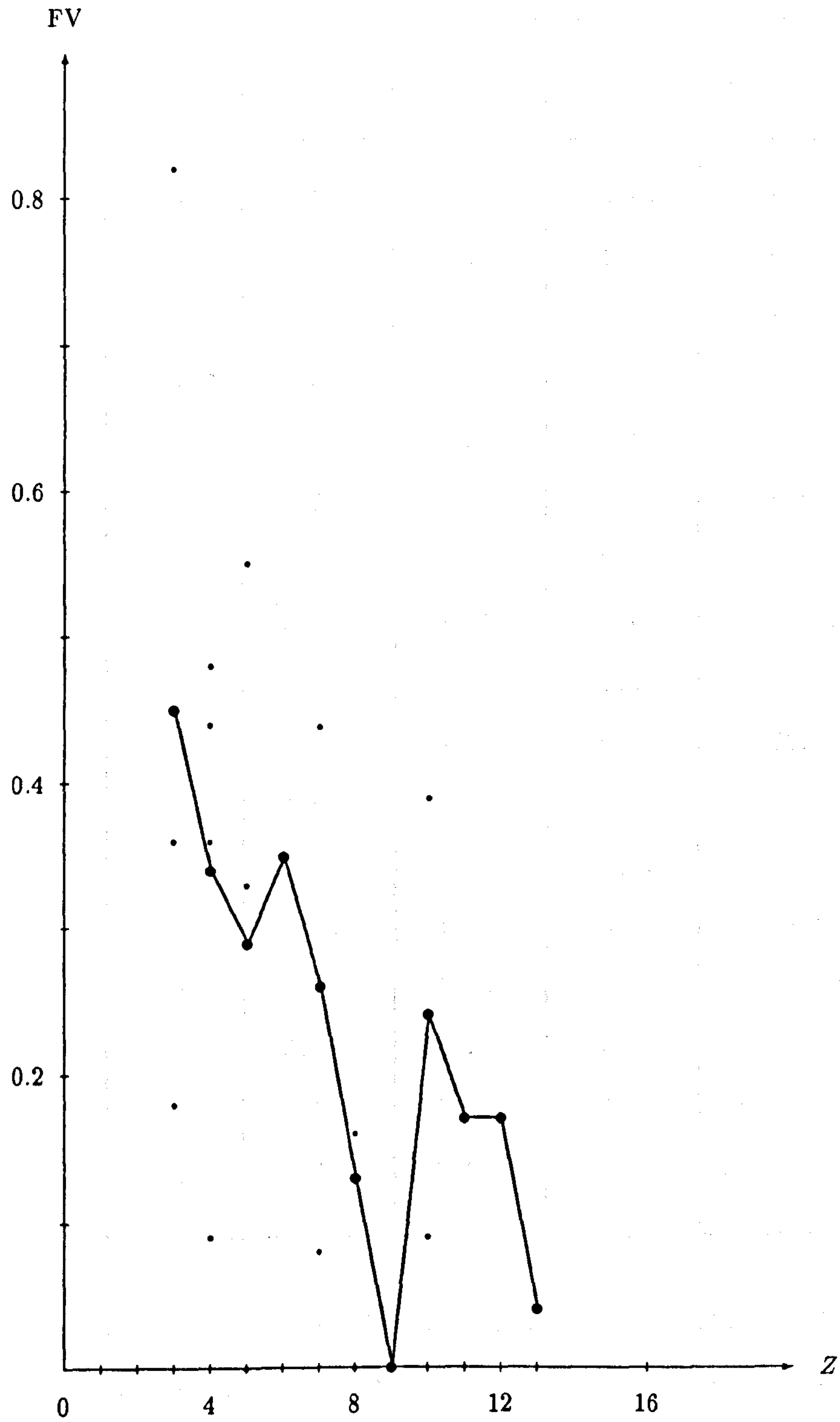

Figure 38

item	FV_1	FV_2	DI	evaluation
1*	0.50	0.66	0.67	easy
1**	0.36	0.50	1.00	acceptable
2*	0.00	0.00	0.00	rejected
2**	0.00	0.00	0.00	rejected
3	0.17	0.25	0.50	difficult
4	0.26	0.17	0.34	difficult
5	0.26	0.25	0.50	difficult
6	0.00	0.00	0.00	rejected
7	0.44	0.34	0.67	difficult
8*	0.17	0.25	0.50	difficult
8**	0.00	0.00	0.00	rejected
9	0.09	0.17	0.00	rejected
10	0.04	0.08	0.17	rejected
11*	0.25	0.25	0.50	difficult
11**	0.09	0.17	0.34	difficult
12*	0.00	0.00	0.00	rejected
12**	0.36	0.42	0.84	acceptable

Table 115. The FV and DI for each question of School 4 ($r = 0.51$). */** for the first/second class only.

item	FV_1	FV_2	DI	evaluation
1a	0.70	0.75	0.50	easy
1b*	0.83	1.00	0.00	rejected
1b**	0.55	0.67	0.67	easy
2a	0.00	0.00	0.00	rejected
2b*	0.33	0.50	1.00	acceptable
2b**	0.09	0.17	0.40	difficult
3a	0.26	0.34	0.67	difficult
3b	0.44	0.42	0.83	acceptable
3c	0.48	0.42	0.83	acceptable
4a	0.91	0.84	0.34	easy
4b	0.52	0.50	0.67	acceptable
4c	0.35	0.17	0.34	difficult
5a	0.35	0.34	0.34	difficult
5b	0.44	0.50	0.67	acceptable
6	0.00	0.00	0.00	rejected
7a	0.57	0.42	0.50	acceptable
7b	0.57	0.50	1.00	acceptable
8a*	0.58	0.50	1.00	acceptable
8a**	0.36	0.17	-0.34	rejected
8b*	0.58	0.84	0.34	easy
8b**	0.00	0.00	0.00	rejected
8c*	0.25	0.25	0.50	difficult
9a	0.61	0.59	0.50	acceptable
9b	0.09	0.17	0.00	rejected
10a	0.57	0.59	0.83	acceptable

10b	0.44	0.50	0.67	acceptable
10c	0.09	0.08	0.17	rejected
11a	0.48	0.50	0.67	acceptable
11b**	0.82	0.84	0.34	easy
$11c_1$**	0.18	0.17	0.34	difficult
$11c_2$**	0.55	0.59	0.50	acceptable
12a*	0.08	0.00	0.00	rejected
12b*	0.00	0.00	0.00	rejected
12T**	0.36	0.42	0.83	acceptable

Table 116. The FV and DI for each item of School 4 ($r = 0.87$). */** for the first/second class only.

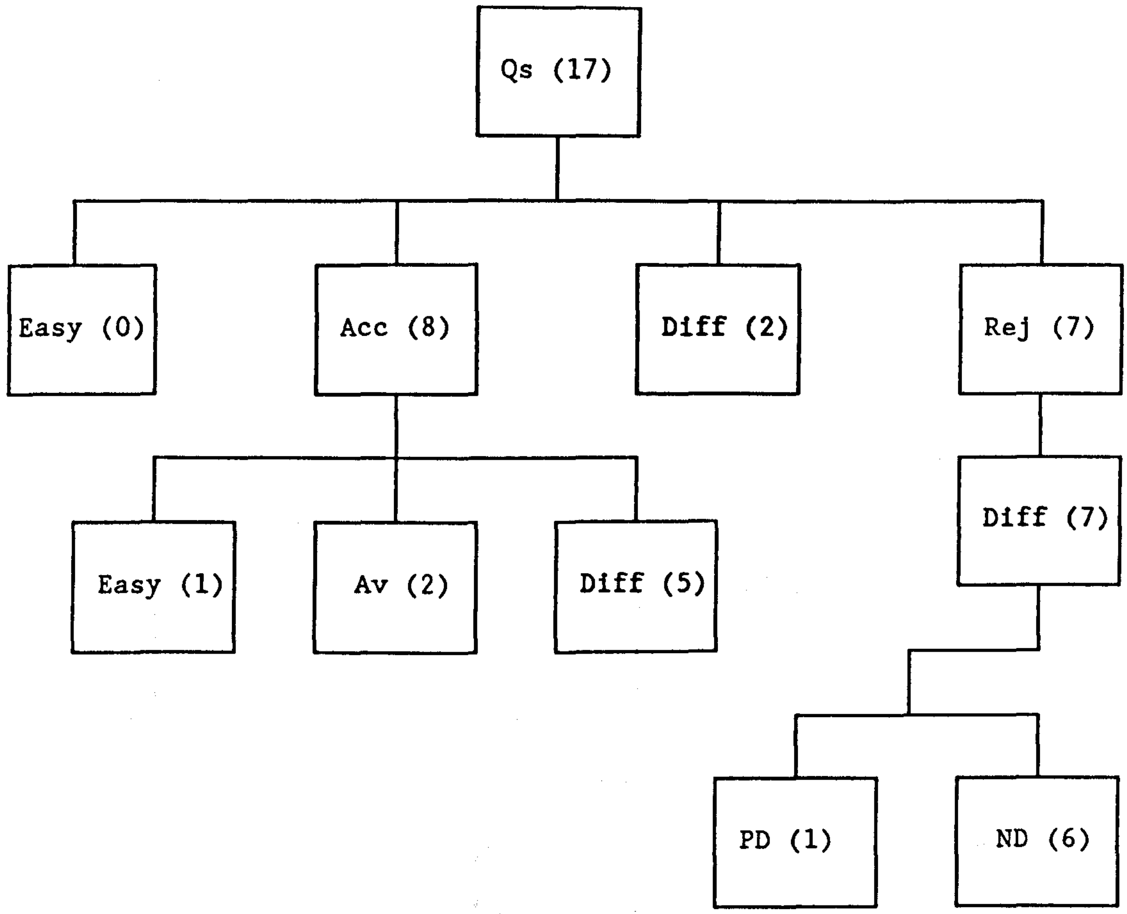

Table 117. The classification of questions' difficulty (School 4).

Key: Acc acceptable, Diff difficult, Rej reject, Av average, PD, ND poor, no discrimination, respectively.

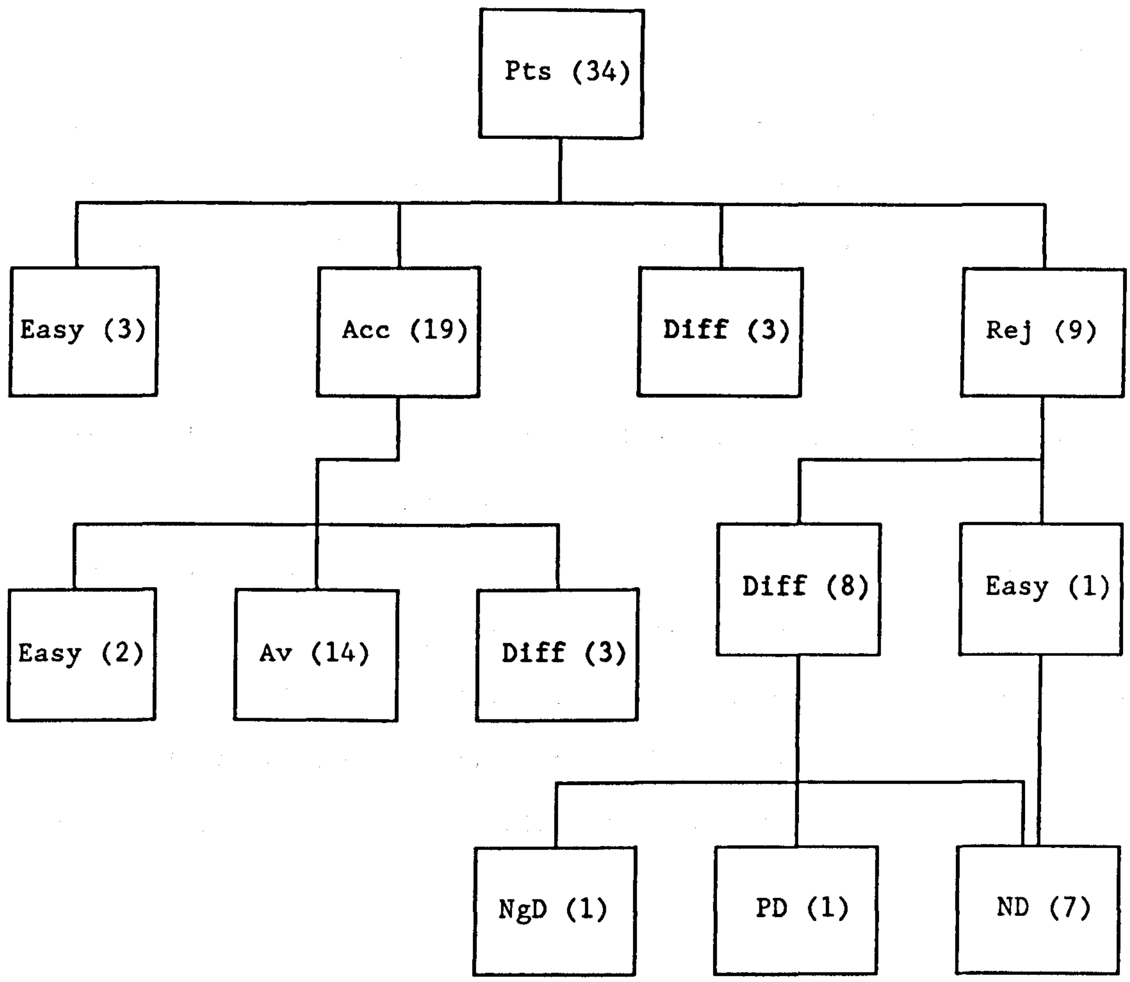

Table 118. The classification of parts' difficulty (School 4).

Key: Acc acceptable, Diff difficult, Rej reject, Av average, ND, PD, NgD no, poor, negative discrimination, respectively.

5. School 5

The test was composed of 9 questions; their classification is given in Table 119. The number of questions which have indivisible, independent and dependent parts is 2, 3 and 4 respectively. The scores are ranked in Table 120 and displayed in Figure 39. The top and the bottom 27% of the sample scored from 53 to 70 (out of 87) and from 11 to 35 respectively.

The FVs of the items and an estimation of their demands are given in Table 121. The average FVs of the items which have the same demand is shown in Table 122 and illustrated in Figure 40.

Tables 123 and 124 give FV, DI, r and the evaluation of items as complete questions and as parts. A significant correlation ($p < 0.01$) between the two FVs was obtained. Tables 125 and 126 show that 6 questions (out of 9) were difficult but 2 of them were acceptable and, of the 15 parts, just 5 were difficult and one of them was acceptable.

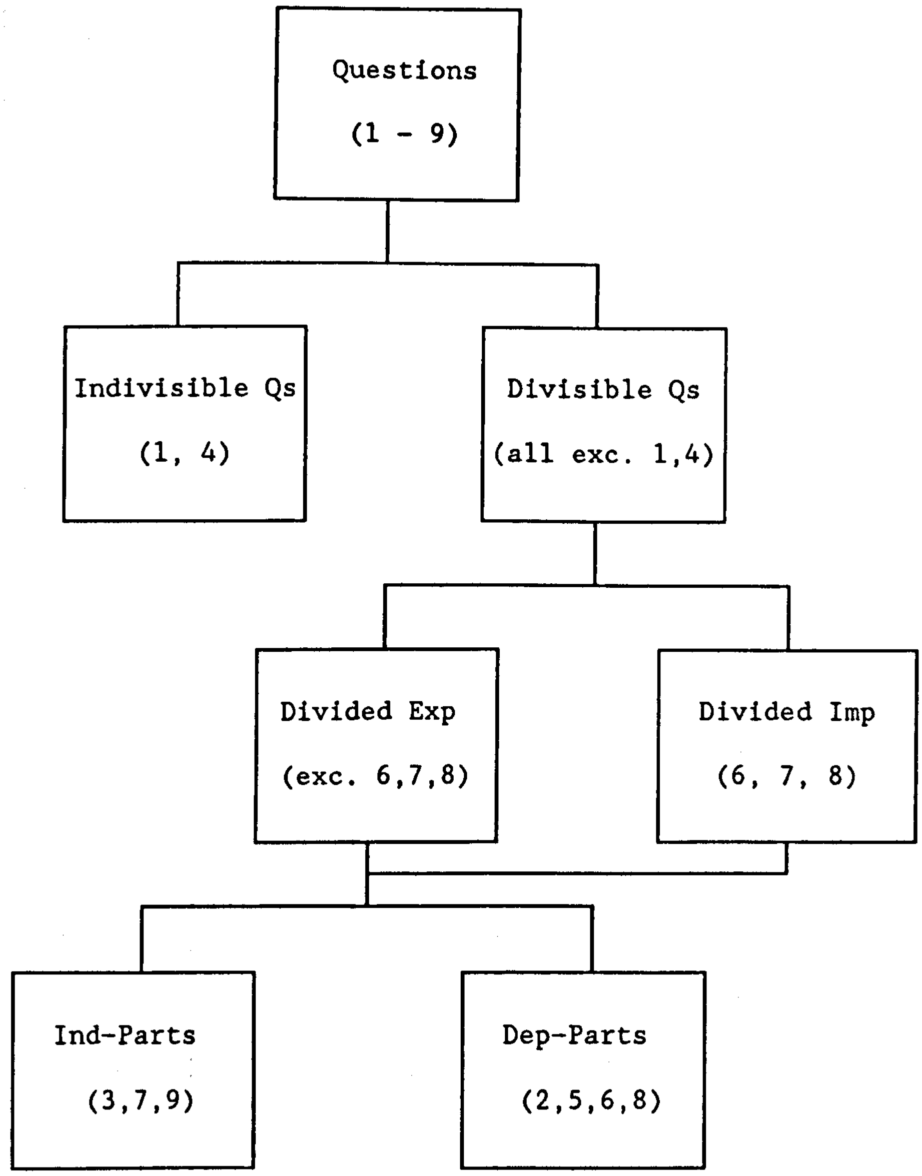

Table 119. The parts' division (School 5).

Key: Exp explicitly, Imp implicitly, Ind independent, Dep dependent.

Frq	Mark
1	70
1	63
1	62
1	61
1	59
2	58
2	56
1	55
3	53
1	49
1	48
1	42
3	41
2	40
4	39
2	38
1	37
3	35
1	32
1	31
2	30
1	28
1	26
1	25
1	24
2	12
1	11

Marks 70–53: ← the top 27% of the sample (N = 13)

Marks 35–11: ← the bottom 27% of the sample (N = 13)

Table 120. The distribution of the test scores of School 5.

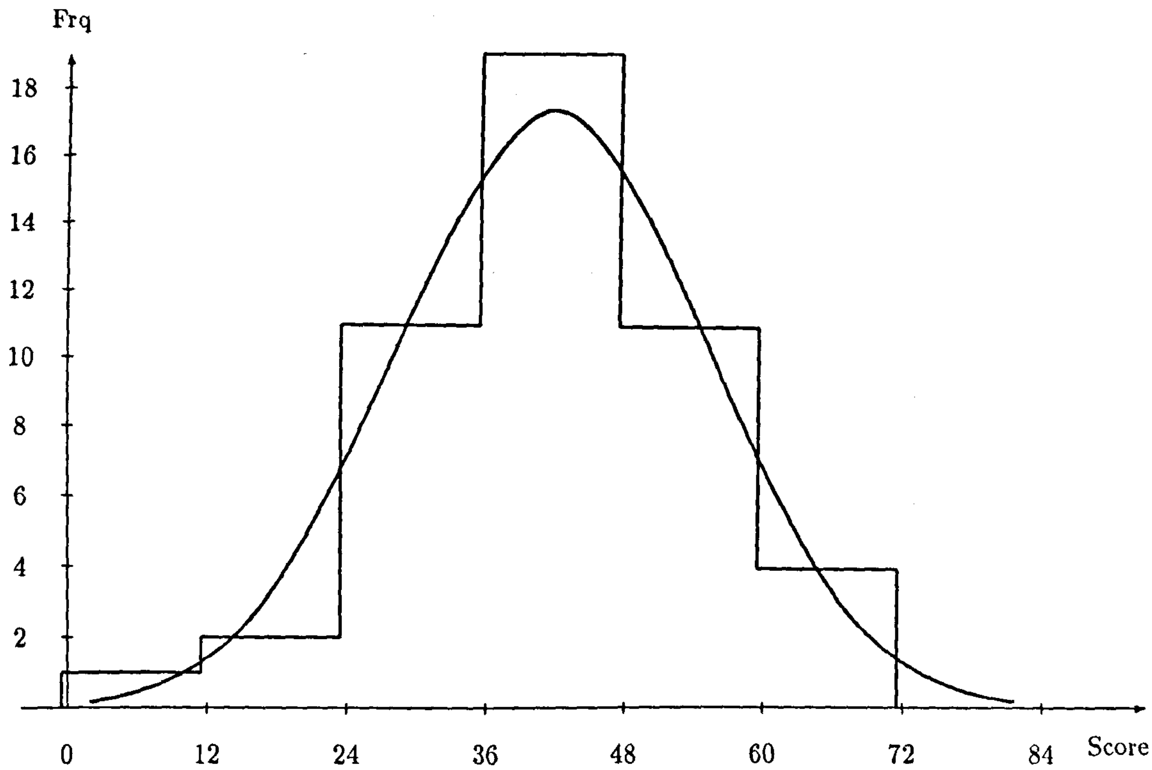

Figure 39

range 59

median 41

mean 41.7

SD 13.3

item		Frq	FV (%)	Z
1		29	60	5
2		18	38	10
3	a	13	27	6
	b	29	60	2
4		31	65	8
5		15	31	13
6		3	6	7
7	1	25	52	4
	2	1	2	11
8		2	4	12
9	a	31	65	5
	b	2	4	9

Table 121. The FV for each item of School 5.

Frq	1	1	2	1	1	1	1	1	1	1	1
FV (%)	60	52	62	27	6	65	4	38	2	4	31
z	2	4	5	6	7	8	9	10	11	12	13

Table 122. The average FVs of items which have the same demand.

item	FV_1	FV_2	DI	evaluation
1	0.60	0.58	0.69	acceptable
2	0.38	0.42	0.69	acceptable
3	0.23	0.27	0.54	difficult
4	0.65	0.58	0.54	acceptable
5	0.31	0.35	0.69	difficult
6	0.06	0.12	0.23	difficult
7	0.02	0.04	0.08	rejected
8	0.04	0.04	0.08	rejected
9	0.02	0.04	0.08	rejected

Table 123. The FV and DI for each question of School 5 ($r = 0.4$).

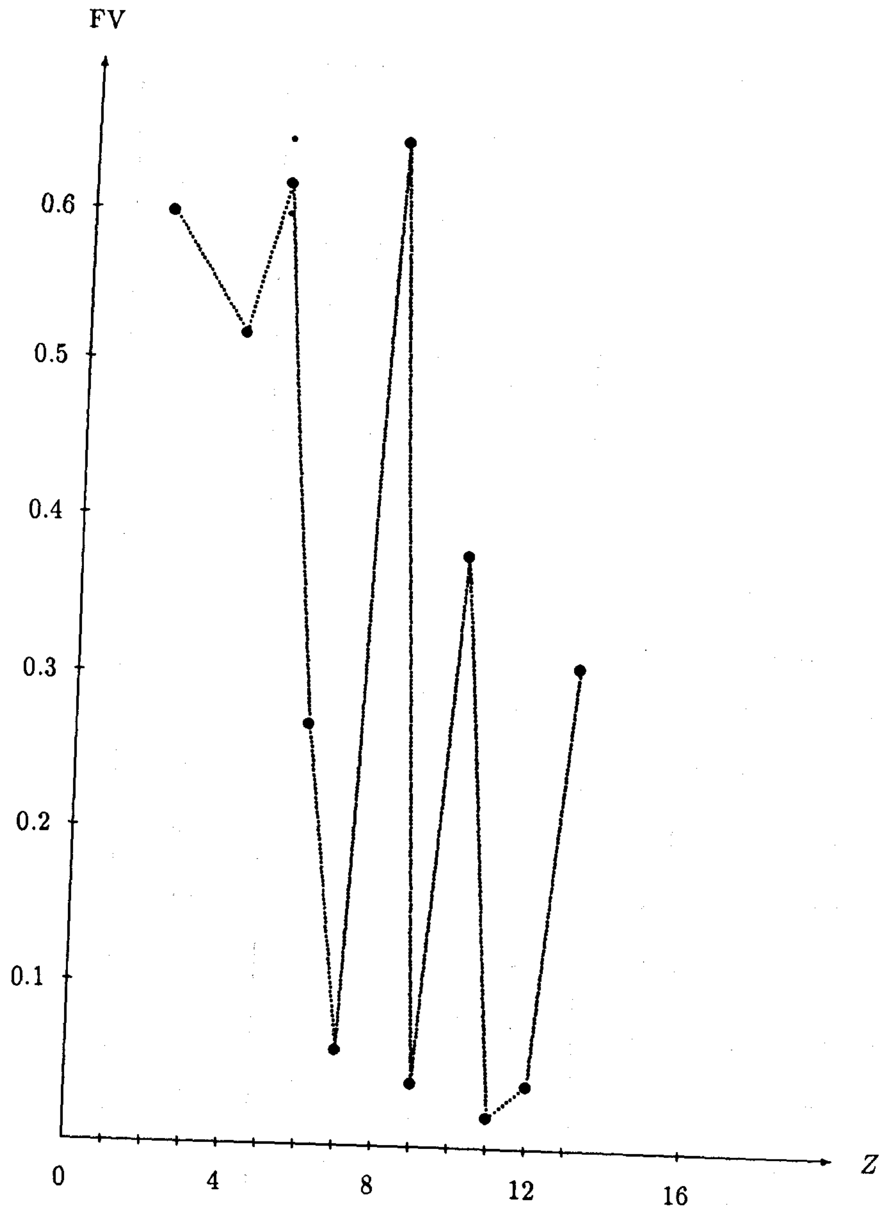

Figure 40

item	FV_1	FV_2	DI	evaluation
1	0.60	0.58	0.69	acceptable
$2i_1$	0.73	0.65	0.54	easy
$2i_2$	0.60	0.65	0.54	easy
2ii	0.58	0.54	0.62	acceptable
3a	0.27	0.27	0.54	difficult
3b	0.60	0.62	0.77	easy
4	0.65	0.58	0.54	acceptable
5a/b	0.54	0.50	0.54	acceptable
5c	0.48	0.54	0.77	acceptable
6	0.06	0.12	0.23	difficult
7_1	0.52	0.58	0.38	improvable
7_2	0.02	0.04	0.08	rejected
8	0.04	0.04	0.08	rejected
9a	0.65	0.77	0.31	easy
9b	0.04	0.08	0.15	rejected

Table 124. The FV and DI for each item of School 5 ($r = 0.64$).

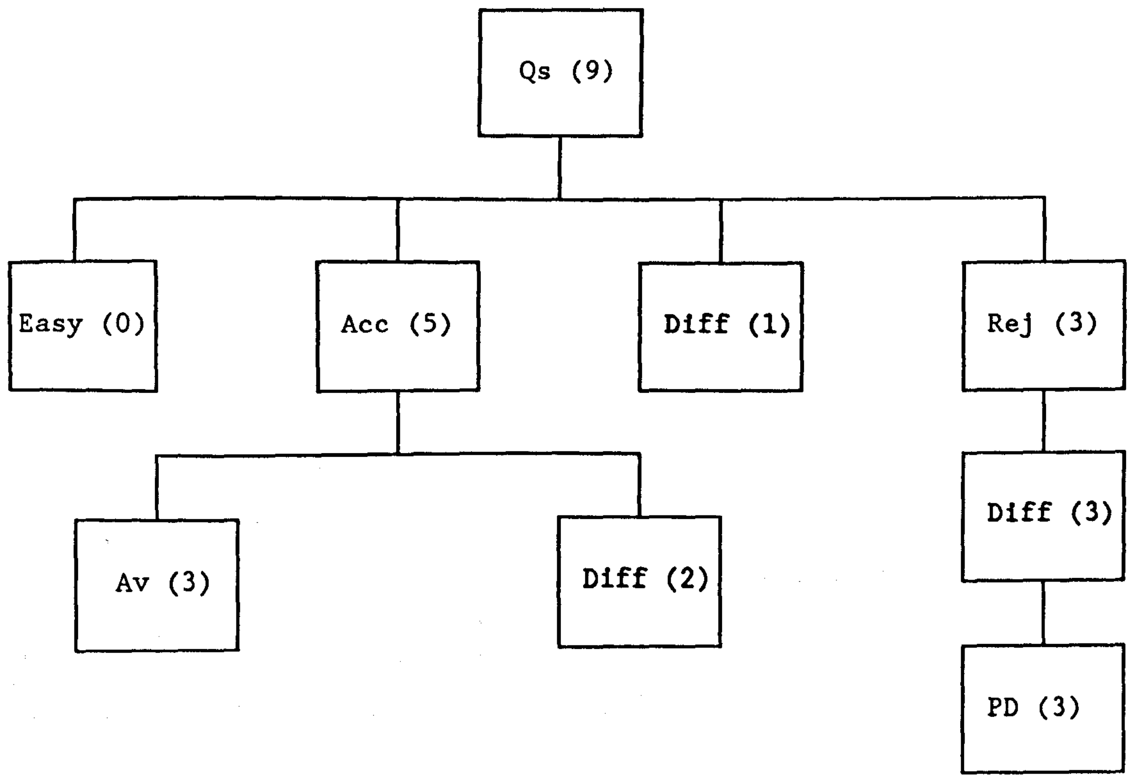

Table 125. The classification of questions' difficulty (School 5).
Key: Acc acceptable, Diff difficult, Rej reject, Av average, PD poor discrimination.

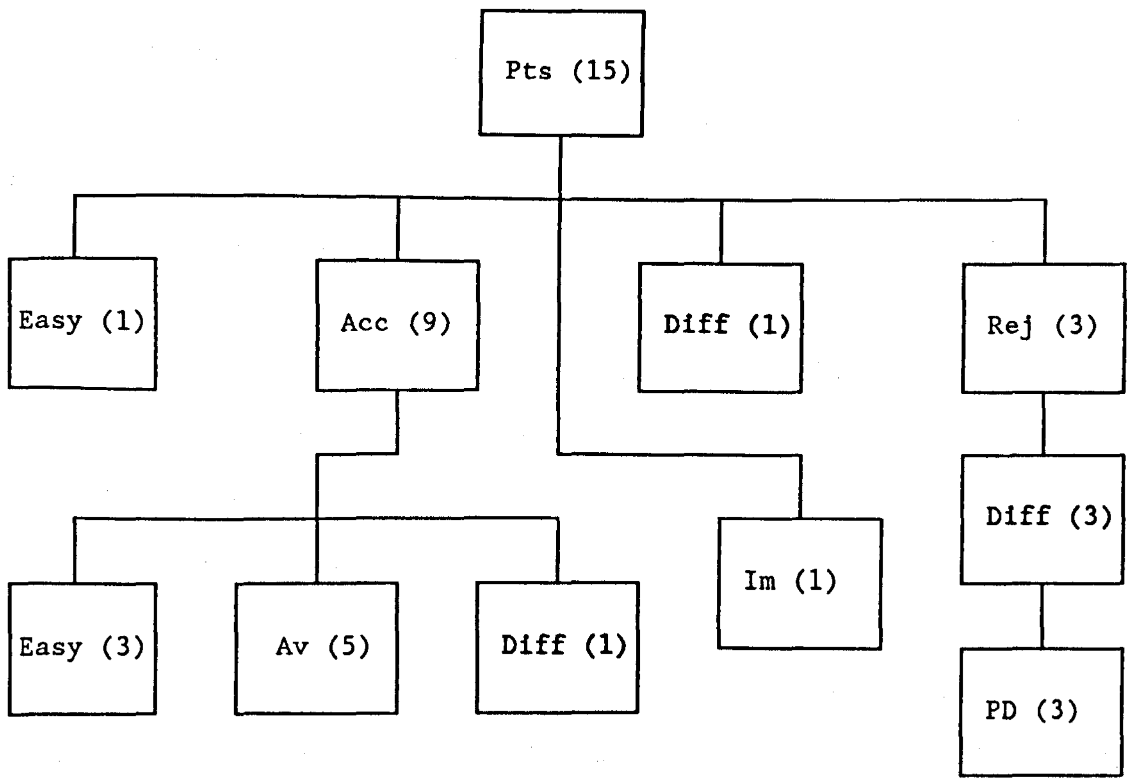

Table 126. The classification of parts' difficulty (School 5).

Key: Acc acceptable, Im improvable, Diff difficult, Rej reject, Av average, PD poor discrimination.

6. The SCE Examination

The same method which was adopted for schools was applied to the SCE Examination sample. The sample size was 121. The test was composed of 13 questions. The number of items which have indivisible, independent and dependent parts is 1, 5 and 7 respectively (Question 9 has an independent part as well as dependent ones and Question 10 is considered to have independent parts). This is recorded in Table 127.

The scores are ranked in Table 128 and displayed in Figure 41. The top and the bottom 27% of the sample scored from 61 to 98 and from 1 to 33 respectively.

The FVs of the items and an estimation of their demands are given in Table 129, whereas Table 130 shows the average FVs of the items which have the same demand. These are illustrated in Figure 42. Once again, the performance in general, decreases when the demand of items increases. Some exceptions (e.g. on the item with $Z = 8$) may due to familiarity.

Tables 131 and 132 give FV, DI, r and the evaluation of items as complete questions and as parts. It was found that (Tables 133 and 134) all questions were difficult but 3 of them were acceptable and that 19 parts (out of 32) were difficult but 9 of them were acceptable. These findings are reflected in the values of reliability coefficient of the test.

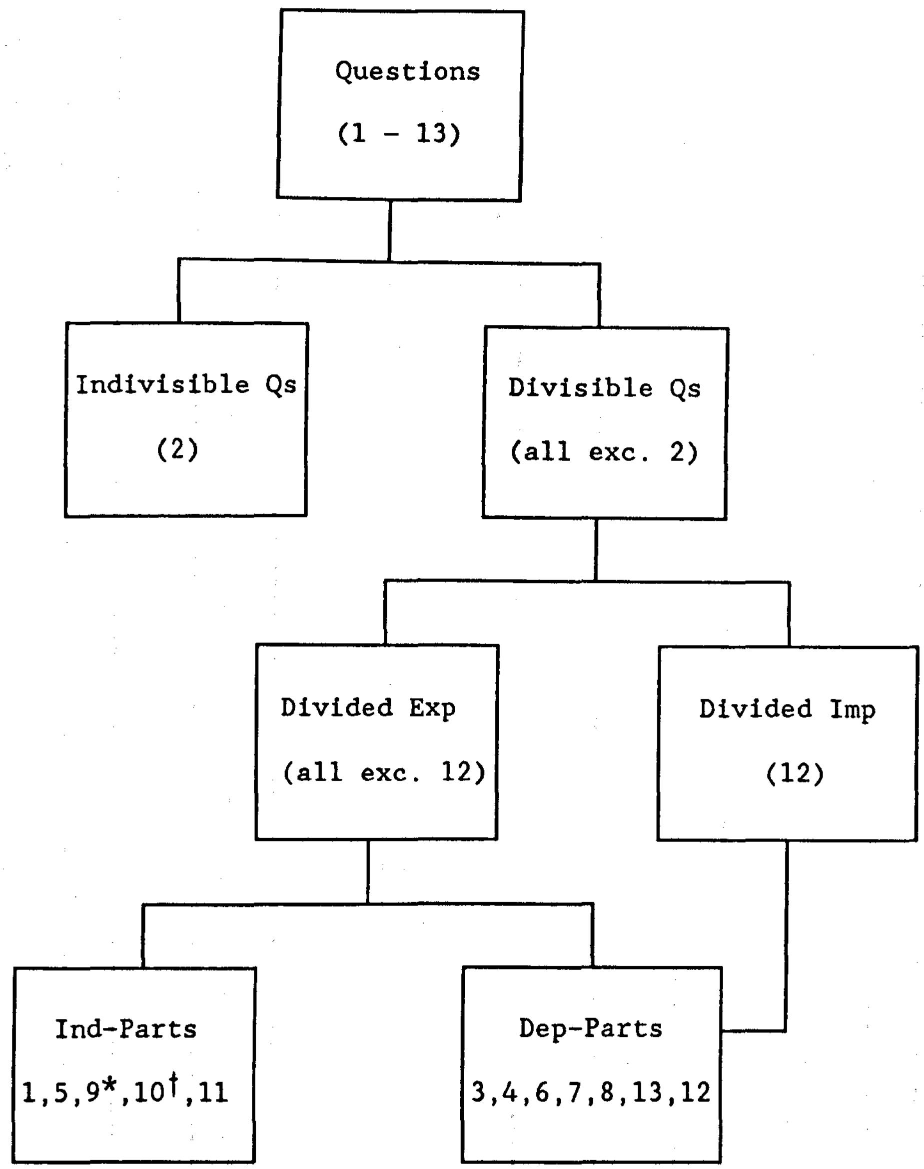

Table 127. The parts' division (SCE).

Key: Exp explicitly, Imp implicitly, Ind independent, Dep dependent. * item has an independent part as well as dependent ones, † item is considered to have independent parts.

Frq	Mark
1	98
1	95
1	88
1	87
1	85
1	84
3	83
1	80
1	77
1	75
1	⟶ 74
2	73
1	72
3	69
2	67
4	66
1	64
3	63
3	62
2	61
4	60
1	59
4	58
3	57
1	56
2	55
2	54
1	51
4	50
2	49
3	48
3	47
2	46
4	45
1	44
1	43
1	42
3	41
3	40
1	39
1	38
1	37
1	36
1	35
3	34
2	33
1	32
3	31
2	30
1	28
1	26
1	25
2	24
1	23
1	⟶ 21
1	20
4	18
1	17
1	14
2	13
2	11
2	10
4	8
2	1

Table 128. The distribution of the test scores (SCE). The top and the bottom 27% of the sample are indicated by ⟹ ($N = 33$ for each of them).

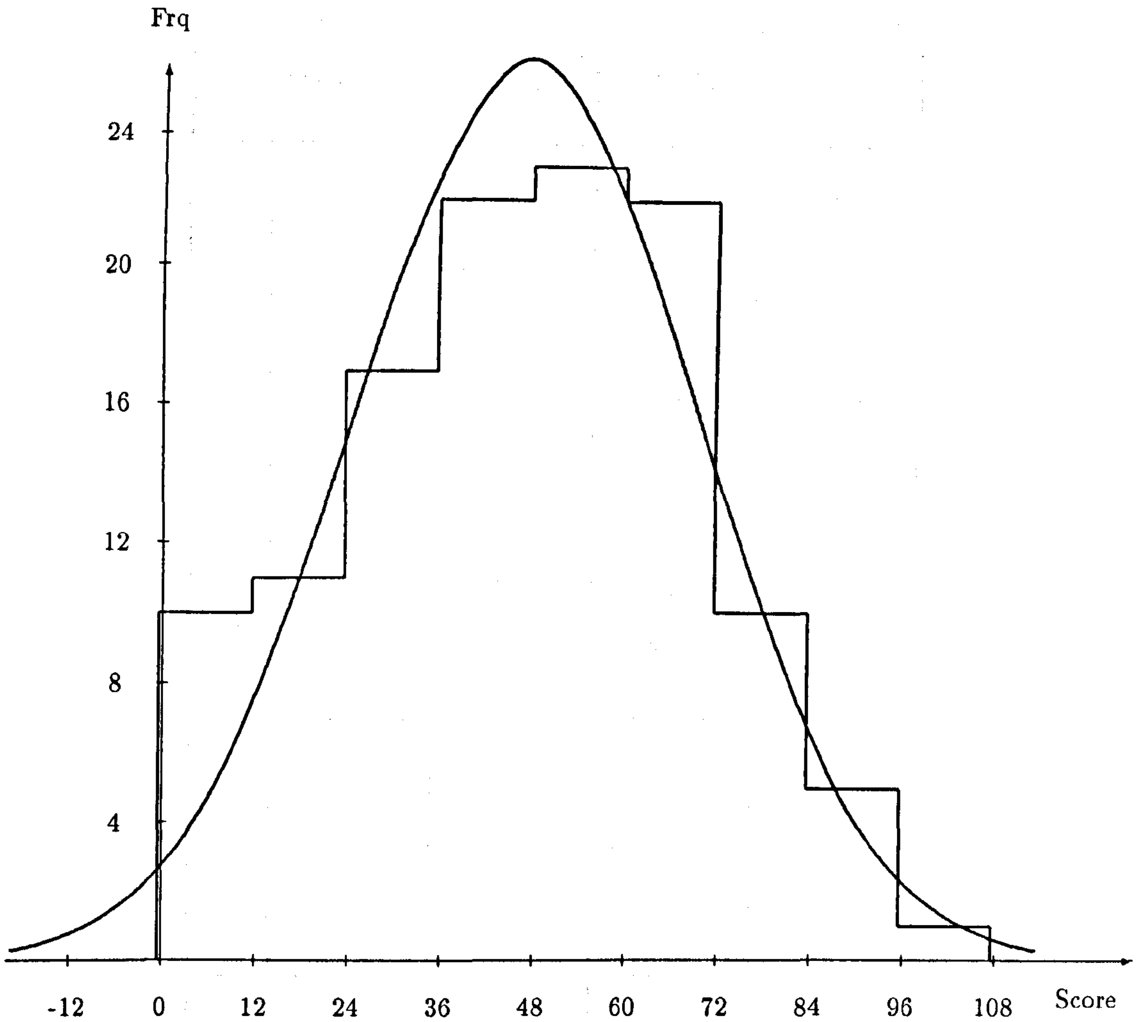

Figure 41

range 97

median 48

mean 46.8

SD 22.1

item		Frq	FV (%)	Z
2		55	46	5
3		24	20	11
4		23	19	12
5	a	72	60	4
	b	40	33	4
6		12	10	11
7		6	5	9
8		6	5	11
9	a & b		82	4
	c	8	7	5
10	a	21	17	5
	b	13	11	6
11	a	10	8	7
	b	9	7	6
12		4	3	10

13		22	18	8

Table 129. The FV for each item of the SCE.

Frq	3	3	2	1	1	1	1	3	1
FV (%)	58	23	9	8	18	5	3	12	19
Z	4	5	6	7	8	9	10	11	12

Table 130. The average FVs of items which have the same demand.

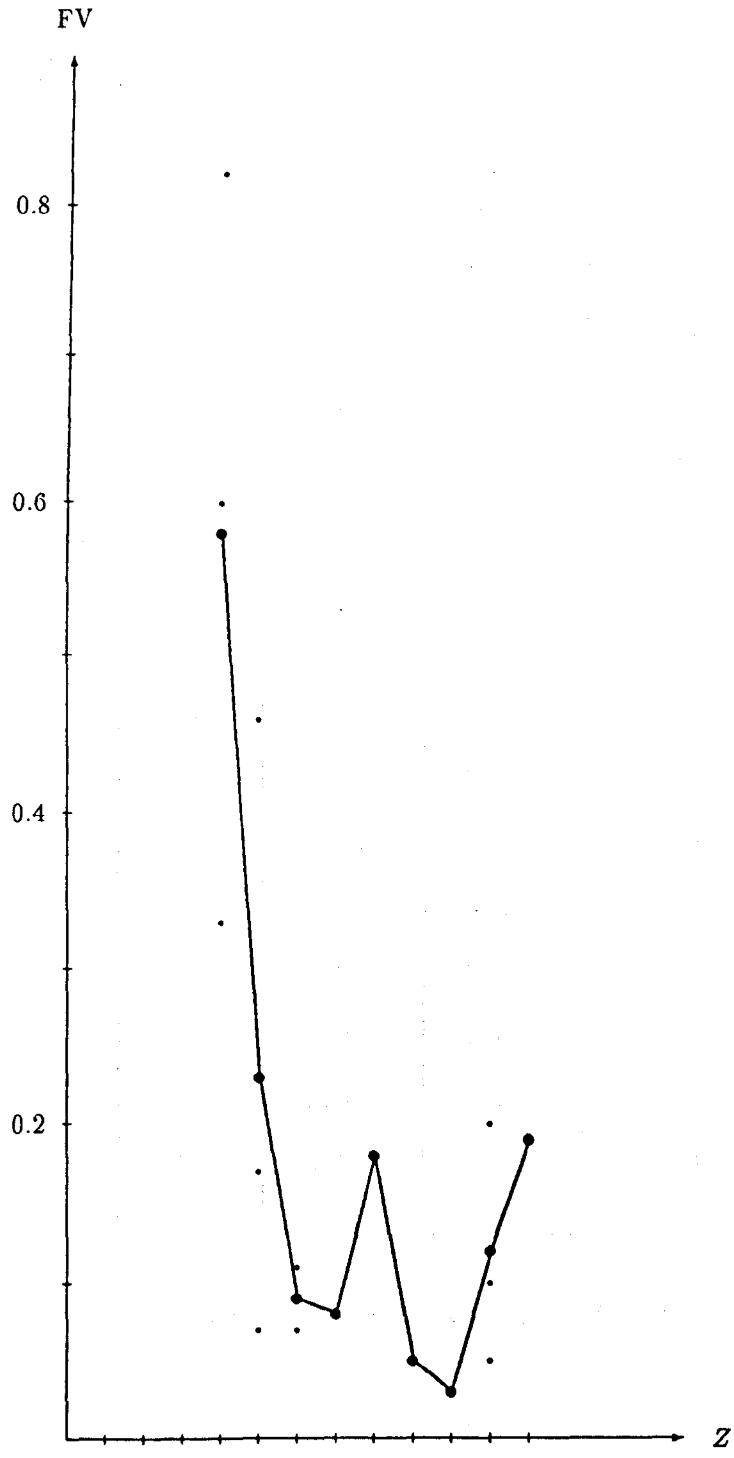

Figure 42

item	FV_1	FV_2	DI	evaluation
1	0.19	0.20	0.39	difficult
2	0.46	0.39	0.61	difficult
3	0.20	0.24	0.36	difficult
4	0.19	0.23	0.39	difficult
5	0.25	0.30	0.61	difficult
6	0.10	0.15	0.30	difficult
7	0.05	0.08	0.15	rejected
8	0.05	0.09	0.18	rejected
9	0.07	0.12	0.24	difficult
10	0.07	0.12	0.24	difficult
11	0.03	0.05	0.09	rejected
12	0.03	0.06	0.12	rejected
13	0.18	0.26	0.52	difficult

Table 131. The FV and DI for each question of the SCE ($r = 0.44$).

item	FV_1	FV_2	DI	evaluation
1a	0.53	0.53	0.70	acceptable
1b	0.25	0.23	0.45	difficult
2	0.46	0.36	0.55	difficult
3a	0.57	0.52	0.91	acceptable
3b	0.32	0.33	0.55	difficult
3c	0.41	0.48	0.42	acceptable
4a	0.40	0.45	0.73	acceptable
4b	0.54	0.56	0.88	acceptable
4c	0.41	0.45	0.79	acceptable
4d	0.23	0.26	0.45	difficult
5a	0.60	0.53	0.76	acceptable
5b	0.33	0.35	0.58	difficult
6a	0.28	0.30	0.61	difficult
6b	0.32	0.33	0.67	difficult
6c	0.27	0.30	0.24	difficult
7a	0.60	0.48	0.55	acceptable
7b	0.05	0.09	0.18	rejected
8a	0.48	0.41	0.70	acceptable
8b	0.45	0.42	0.48	acceptable
8c	0.12	0.18	0.36	difficult
8d	0.12	0.17	0.27	difficult
9a	0.84	0.80	0.39	easy
9b	0.89	0.85	0.30	easy
9c	0.07	0.12	0.24	difficult
10a	0.17	0.27	0.55	difficult

10b	0.11	0.14	0.27	difficult
11a	0.08	0.15	0.24	difficult
11b	0.07	0.14	0.27	difficult
12_1	0.12	0.20	0.39	difficult
12_2	0.03	0.06	0.12	rejected
13a	0.34	0.41	0.64	acceptable
13b	0.22	0.29	0.52	difficult

Table 132. The FV and DI for each item of the SCE (r = 0.88).

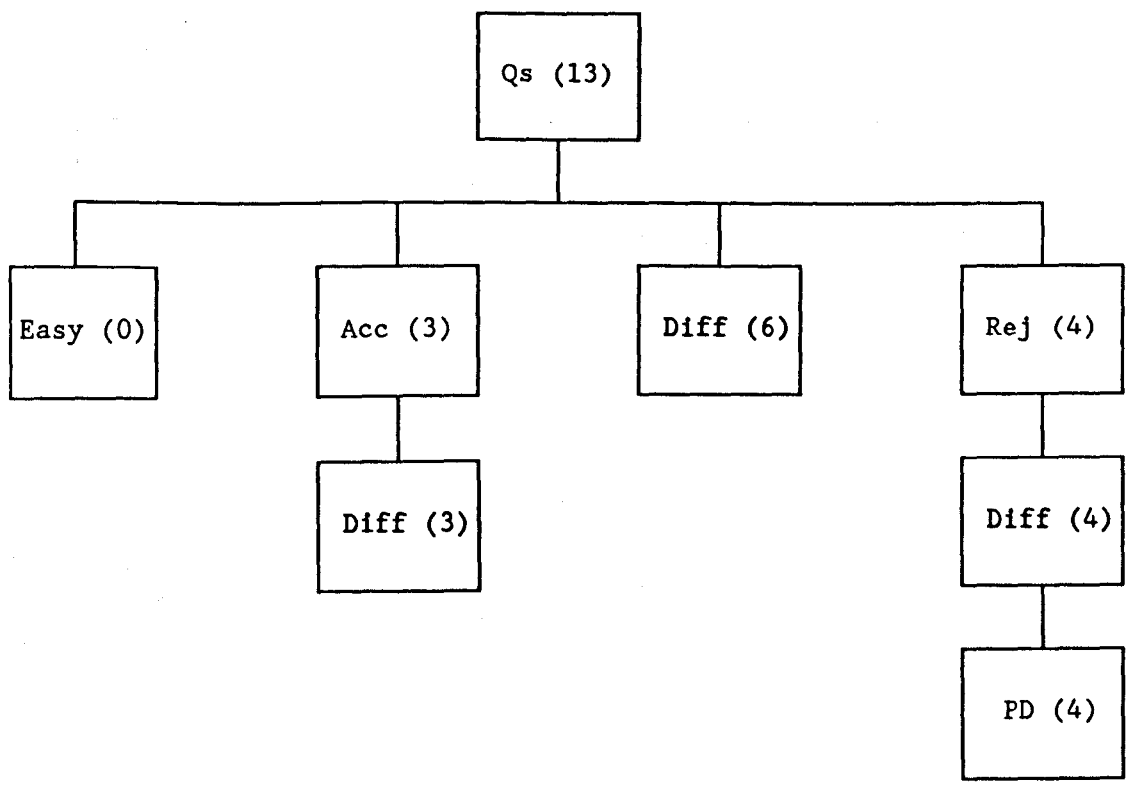

Table 133. The classification of the questions' difficulty (SCE). Key: Acc acceptable, Diff difficult, Rej reject, PD poor discrimination.

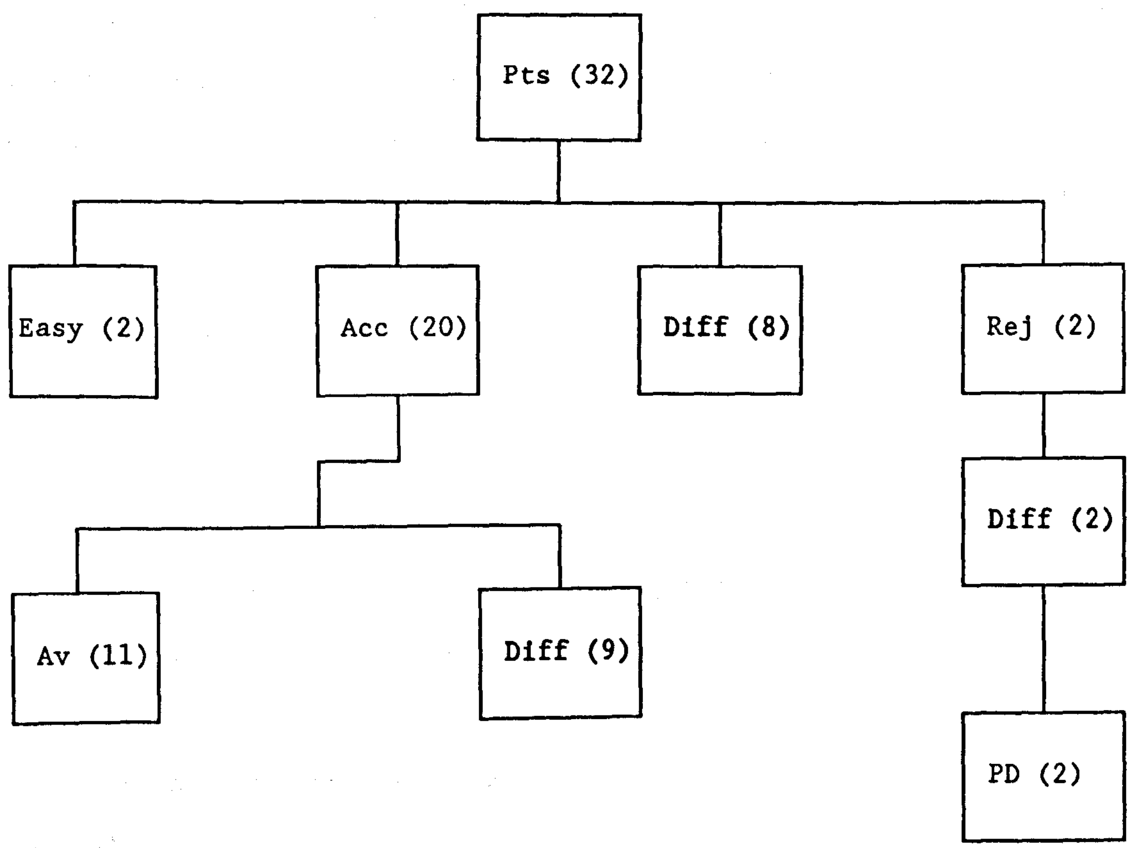

Table 134. The classification of the parts' difficulty (SCE).

Key: Acc acceptable, Diff difficult, Rej reject, Av average, PD poor discrimination.

Study of some difficult items

In order to find out the reasons for difficulties of items, an analysis of some parts (chosen with a FV below 0.4) has been made for each school and the SCE Examination. The difficulty of a question may arise from just one part and the reason for this difficulty may come from the subject matter or other factors (such as language, structure, type, appearance, etc.) or any combination of them.

For each of the chosen items, The percentages of pupils gaining the various marks are listed together with the main mathematics concepts involved in the item and some other factors which may have caused difficulties.

Items 9(b) and 8(c) from School 1

For the first item, the percentages getting 5 (out of 5), 4, 3, 2, 1, 0 and not attempting were 3, 0, 3, 0, 21, 64 and 9 respectively. Therefore, 73% (i.e. 64% + 9%) of pupils had no marks. The mathematics concepts required are components of a vector, scalar product. The diagram in this item was inaccurate (not an equilateral triangle, AD appears to be perpendicular to BC). Some information is given out of order (D is used before it is defined).

For the second item, the percentages getting 5 (out of 5), 4, 3, 2, 1, 0 and not attempting were 9, 12, 20, 15, 23, 9 and 12 respectively. So 21% of pupils had no marks. The mathematics concepts required are quadratic equation, the discriminant and roots of a quadratic. The existence of a parameter in the equation and the negative form of the question may be some of the causes of the difficulty of this item.

Item 10 from School 2

The percentages getting 5 (out of 5), 4, 3, 2, 1, 0 and not attempting are 0, 4, 4, 20, 52 and 20 respectively. Therefore, 72% of pupils had no marks. The mathematics concepts involved are the same as in the previous item. The abstract form of this item (the equation has one variable and three parameters) and the interpretation of "the nature of roots" may be the reasons for its difficulty.

Items 9(b), 10(i) & (ii) from School 3

For the first item, the percentages getting 5 (out of 5), 4, 3, 2, 1, 0 and not attempting are 0, 1, 1, 12, 2, 56 and 28 respectively. So, 84% of pupils had no marks. The mathematics concept required is the sum of the first n positive integers. This item is in an abstract form, the pupils need an appropriate strategy such as "break down" (breaking the problem into smaller steps) to solve it and it seems to me that this item is unfamiliar.

For the second item, the percentages getting 3 (out of 3), 2, 1, 0 and not attempting are 11, 1, 4, 61 and 23 respectively. Therefore, 84% of pupils had no marks. The mathematics concept involved is similarity between two triangles. The diagram is not accurate (incorrectly proportioned), A has two meanings and there is irrelevent information. This and the need for strategy may raise the difficulty of this item.

For the third item, the percentages getting 4 (out of 4), 3, 2, 1, 0 and not attempting are 14, 8, 6, 22, 26 and 24 respectively. So 50% of pupils had no marks. The mathematics concepts required are area of a rectangle and maximum of a function. The interpretation of "greatest area" and the concept of maximum itself may be some reasons for the difficulty of this item.

<u>Item 12(b)* from School 4</u>

The percentages getting 6 (out of 6), 5, 4, 3, 2, 1, 0 are 0, 8, 0, 8, 8, 0 and 76 respectively. So 76 % of pupils had no marks. The mathematical requirements are the formulae for tan 2θ, sin 2θ and cos 2θ. It seems to me that faulty recall of trigonometric formulae and the complex reasoning required to factorise and solve the resultant trigonometric equation are the reasons for the difficulty of this item.

<u>Item 9(b) from School 5</u>

The percentages getting 7 (out of 7), 6, 5, 4, 3, 2, 1, 0 and not attempting are 4, 0, 0, 2, 4, 4, 34, 46 and 6 respectively. Therefore, 52% of pupils had no marks. The mathematics concepts involved are position vector of a point, parallelogram and collinearity of three points. Although this item is already broken into steps, the abstract form of both notation and ideas may cause its difficulty.

<u>Items 12_2, 12_1, 7(b) and 10(b) & (a) from the SCE Examination</u>

For the first item, the percentages getting 4 (out of 4), 3, 2, 1, 0 and not attempting are 3, 1, 1, 8, 16 and 71 respectively. So, 87% of pupils had no marks. The mathematics concepts required are parallelogram, position vector and collinearity of three points. As we noted in the previous item, the abstract notation and ideas may be the reasons for the difficulty.

For the second item, the percentages getting 3 (out of 3), 2, 1, 0 and not attempting are 12, 9, 5, 61 and 13 respectively. So, 74% of pupils had no marks. The mathematics concept involved is the position vector of a point. The same remarks as for the previous item may be valid here.

For the third item, the percentages getting 4 (out of 4), 3, 2, 1, 0 and not attempting are 5, 44, 9, 7, 23 and 12 respectively. Therefore, 35% of pupils had no marks. The mathematics concepts required are components of a vector, scalar product and the image of a point under reflection. Maybe the three dimensional aspect of this item is one reason for its difficulty.

For the fourth item, the percentages getting 4 (out of 4), 3, 2, 1, 0 and not attempting are 11, 13, 12, 24, 8 and 32 respectively. Therefore, 40% of pupils had no marks. The mathematics concepts involved are differentiation and maximum of a function. Failure to realise that this item can be solved without having completed the previous part and failure to check that the value of x obtained does indeed give a maximum value of V are two reasons for the poor performance.

From the fifth item, the percentages getting 3 (out of 3), 2, 1, 0 and not attempting are 17, 6, 8, 60 and 9 respectively. So 69% of pupils had no marks. The mathematics concepts required are area of a side of rectangular box and its volume. The interpretation of the data and the three dimensional aspect may be some reasons for the difficulty of this item.

A comparison between Paper II of schools and the SCE Examination

(i) Mean score

Table 135 gives the mean scores and standard deviations for Paper II of the schools and the SCE Examination. The average of the means from the schools is significantly different ($p < 0.05$) from the mean score from the SCE. This may be explained by the fact that the SCE examination was, in general, harder than those of schools since it came later then the internal exams.

(ii) Reliability coefficient

Table 136 gives the reliability coefficient for Paper II of the schools and the SCE Examination (for both questions and parts). There is no significant difference between the average reliability coefficient for the school examinations and the reliability coefficient for the SCE Examination in the case of whole questions but the difference is significant ($p < 0.1$) in the case of parts.

sample	mean*	SD
SCE	46.81	22.19
Sch 1	53.00	17.61
Sch 2	54.62	20.89
Sch 3	47.83	21.55
Sch 4	57.72	19.43
Sch 5	47.92	15.17

Table 135. The mean scores and standard deviations for Paper II.

* possible score is 100.

sample	reliability coefficient	
	questions	parts
SCE	0.44	0.88
Sch 1	0.07	0.80
Sch 2	0.40	0.78
Sch 3	0.51	0.86
Sch 4	0.68	0.87
Sch 5	0.40	0.64

Table 136. The reliabilities coefficients for Paper II.