An Investigation into the Statistical Understanding of 12-18 Year Olds

by

DAVID J. CASSELL, B.Sc.(Tech) Statistics, P.G.C.E.,
Postgraduate Diploma in Statistics & Statistical Education

A thesis submitted in partial fulfilment of the requirements of Sheffield Hallam University for the degree of Master of Philosophy

May 1993

TABLE OF CONTENTS

Abstract

The aim of this project was to develop a model for the structure of the development of statistical thinking in students of secondary school age, i.e. 12-18. Previous research has tended to concentrate on individual problems and no large scale research has been carried out in this area. The aim was therefore to produce a model which encompassed all areas of Statistics and showed the building up of concepts.

The basis of the model was a hierarchical structure based on Gagné's Cumulative Learning Theory, with due allowances made for subsequent criticisms of the rigidity of such a model. Models were proposed in five areas considered to involve the main principles of elementary statistics. Superimposed on to these maps of conceptual development was a 3-stage structure corresponding to classical Piagetian stages.

Prior to testing a detailed survey was made of available techniques for examining the validity of such models. In particular the Inclusion Analysis technique devised by Clarke & White was carefully examined noting cases where it was inappropiate or invalid.

After some initial testing and expert analysis the initial models were modified. The strength of the restructured models was examined by presenting detailed written tests to over 200 students in the age range under investigation. Using Clarke & White Inclusion Tests the strength of links between the concepts was tested and some justification given to the ordering of concepts in the hierarchy and adjustments made where necessary. The validity of grouping skills into 3 stages was tested and an attempt made to correspond these to age using correlation techniques.

Although, from the data collected the full detail of the model could not be entirely supported, there was evidence to justify the main framework and certain key linkages to produce a final model. This enabled a detailed analysis of the National Curriculum and its United States counterpart to take place in terms of age-related content and structure. Suggestions were also presented to writers and curriculum designers in the light of research findings.

Acknowledgements

In submitting this thesis I would wish to express my gratitude to the following:-

Professor Warren Gilchrist - as tutor and mentor of this project throughout, it would not have been completed without his help.
Dr. Stuart Trickey - his tremendous enthusiasm and thorough assistance has improved the quality of this project enormously.
Mrs Margaret Rangecroft - her initial inspirations and continued help and support throughout have been an invaluable assistance.
The Master & Fellows of Magdalene College, Cambridge - the generosity of the Fellowship which I was granted in the Michaelmas Term of 1991, the study facilities with which I was provided, and the interest and encouragement which was shown towards this project made the completion of it possible.
Mr Chris Wade, Headmaster of Hewett School, Norwich, whose efforts made the above Fellowship possible.
The Staff & pupils of the Hewett School, Norwich - the help given in trialling and testing and commentary on material was invaluable.
David Green of Loughborough University & Peter Holmes and other staff at Sheffield Hallam University who assisted in the expert analysis.

Introduction

The inspiration for this project came as a student on a Diploma course where the researcher was required to complete an assignment on the development of a concept in a particular area of Probability/Statistics. The concept of the mean was chosen as it was expected that there would be a great deal of research in this area. It came as a great surprise therefore to find that not only had there been very little consolidated research in the understanding of this concept but in the development of statistical concepts as a whole.

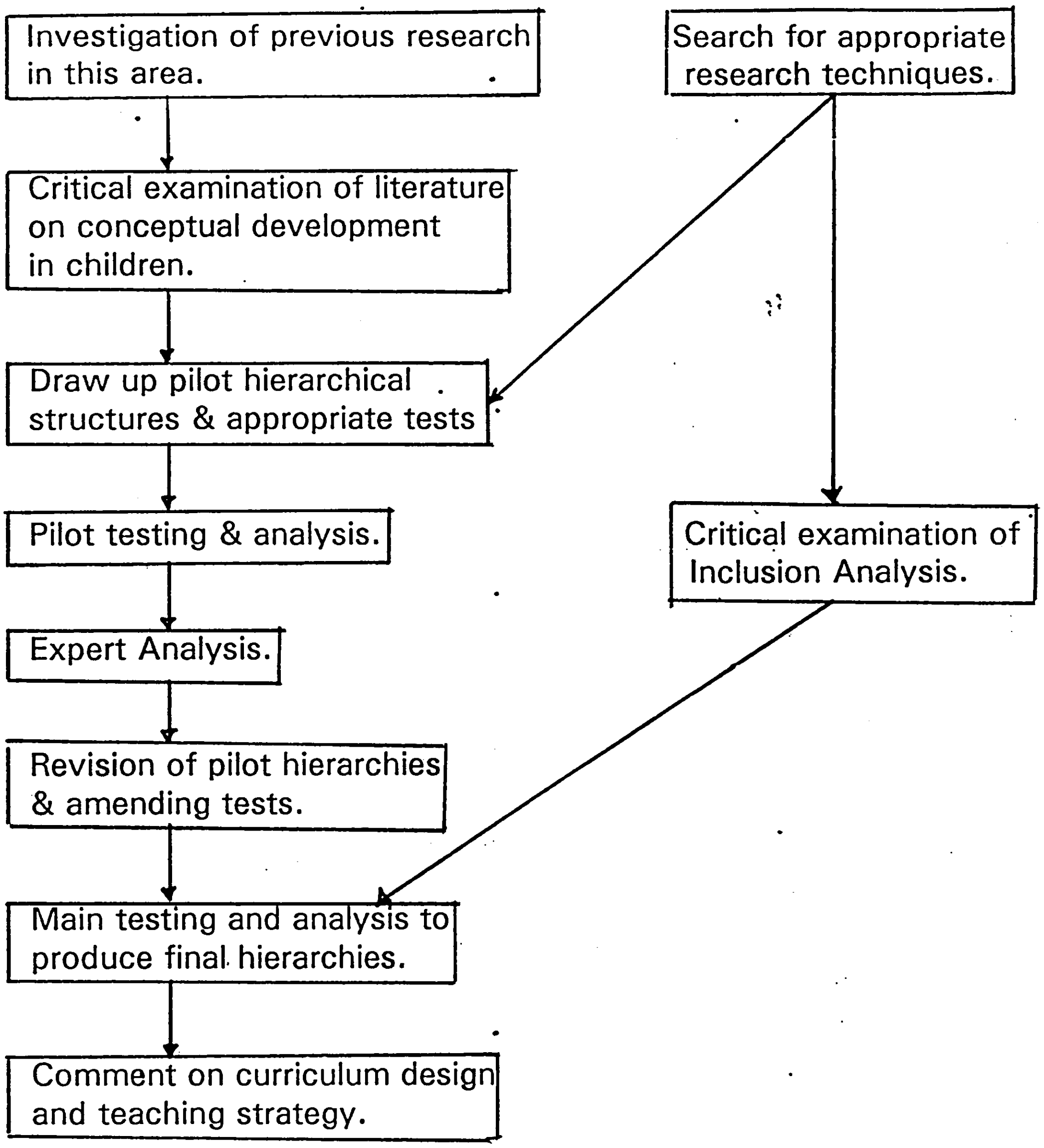
The lack of research, and hence knowledge, in this area was particularly disturbing in view of the large inclusion of 'Data Handling' in the new National Curriculum. Having been an area previously neglected in traditional school syllabuses past experience is limited. The question arose as to the basis on which those who had composed the National Curriculum and those who would be writing accompanying teaching material could judge appropriate levels and ordering. The National Curriculum gave specific ages by which on average topics were expected to be known, and the order in which these should be approached. Were these age expectations reasonable and did the proposed ordering of topics facilitate a good understanding of the basic concepts? Similarly on what basis are the producers of teaching material to make judgements on topic orderings and teaching styles?

The purpose of this project was then to begin to establish a framework to describe the way in which children's understanding in the field of Statistics should be developed. What were the natural intuitions which would either aid or hinder the learning process? What was the 'normal' pattern of the development of statistical concepts and at what age could children be expected to have developed the mental capacity to properly understand these concepts?

The early work in this project is concerned with looking at previous work carried out both in statistical concepts and in more general cognitive development. The purpose of this was to suggest how a framework for the development of statistical concepts could be set up. Once a framework was established this had to be rigorously tested. This presented problems as there appeared to be no widely used techniques available and the technique used had to be examined itself for validity. Only when the structure of the development of concepts had been established could the National Curriculum be critically examined and suggestions made as to suitable teaching styles.

The following flow diagram shows the progression of work.

Flow Diagram Showing The Pattern of Research in This Thesis



Chapter 1

Research into Children's Understanding of Statistical Concepts

1.1 Introduction

At the outset of this project there appeared to be lttle consolidated research in the area of statistical concepts. An extensive literature search revealed a number of disjointed attempts to investigate this area with no particular pattern. However, over the period during which the project was carried out, several new sources of material were published which, whilst not having a significant influence, were relevant to the work in this project. Indeed contact was made with the International Study Group on the Teaching of Statistics which provided many new sources. There has been great interest in the last few years in the development of understanding of statistics in children and it is hoped this project will make a significant contribution.

In the area of probability (i.e. chance) there is a long history of research from the early studies of Piaget to the more recent detailed investigation of Green[1982]. The only other area in which there appears to have been any significant study into statistical concepts is incorporated in various studies of graphical interpretation which have included statistical diagrams. Where statistical understanding has been tested this has largely been in the area of measures of location (i.e. mean, median & mode), and a little on measures of spread with a few miscellaneous topics such as areas under a normal curve. Other ideas such as correlation and even simple understanding of tables appear to have been given little attention.(Kahneman, Slovic, Tversky [1982])

The research so far carried out in statistics appears to fall into two main categories:-

(a) Large scale surveys of general mathematical ability using a small number of statistical items . Although the questions are limited in number and the depth to which they examine statistical concepts is in some cases doubtful, these projects nevertheless have been carried out on large numbers of school children and give useful insights into the area of research.

(b) Small scale surveys carried out on limited samples, often college students, to examine specific areas of concern in more detail. Some are designed to examine the reason for a particular misconception. These are often used as pre- & post-tests for courses in certain topics. Although giving more detailed information on certain topic areas the results are not useful for general interpretation as the samples are small and usually represent a limited ability and age range.

1.2 Large Scale Surveys

Amongst the large scale surveys mentioned above the most notable in the UK are those of the Assessment & Performance Unit (APU) [1981, 1987] and the National Foundation for Educational Research (Cresswell & Grubb, 1987). Each of these surveys was carried out on large numbers of 11-16 year olds. The tests contained a small number of questions examining:-

(i) understanding & ability to draw statistical diagrams

(ii) ability to calculate simple location measures.

With regard to statistical diagrams, graphs, block/bar graphs and pie charts were used. The Cambridge Trust of Education(CTE) [1981] in examining APU findings reported that whilst the interpretation of diagrams was generally well understood by even the younger children, the level of success was dependent on the nature of the scale used and in particular whether interpolation was required. Construction of diagrams proved much more difficult and the APU reported [1986] that drawing graphs, histograms, cumulative frequency and pie charts was successful in less than 50% of cases. The CTE reported similarly that constructing and completing graphs proved difficult and that pie charts were the most difficult. Care clearly needs to be taken in interpreting these results. If pupils are able to make inferences from diagrams this indicates that a basic understanding of the nature of the information portrayed exists. The difficulties of scale interpretation are probably due to a lack of understanding of scales rather than the statistical concept involved. Further, the construction problems could result from a number of other difficulties such as the inability to use simple equipment e.g. a protractor.

The main area of statistical interest however in these tests is the recognition and calculation of the main measures of location, i.e. the mean, median and the mode. Findings seem to differ slightly with the CTE & APU reporting that the success rate in calculating these was about 50%, but Cresswell & Grubb[1987] found an 80% success rate on calculating means (the only measure tested with secondary pupils). Several comments however are fairly general. Firstly, all report a better response to questions on the mean using the word 'average' and in particular the term 'arithmetic mean' seems to have caused confusion. APU further report that many teachers administering the test commented on the language difficulties caused for some less able pupils. They also report only half the subjects recognized the definitions, but this could again be due to language difficulties. The CTE findings tend to indicate that most subjects were aware of the different measures and that there was no confusion between these. Cresswell & Grubb noticed that some subjects gave the total when asked for the mean and that this became more prevalent when non-integer values were used for data. None of the questions appearing in these tests seems to have examined the appropriateness of any of the measures in different situations.

Other items in the NFER survey asked pupils to comment on how they would carry out a survey and on areas under the normal curve. In the area of survey techniques the main difficulty appears to be in assuming that a random method necessarily gave an unbiased sample, and some failed to see the need for a sample to be unbiased; that is they thought that a biased sample was a good thing. Questions on normal curves gave a poor response (38%) and teachers commented that this was an area pupils were not familiar with.

1.3 Large Scale Research in the United States

Similar testing to the above has been carried out by the National Council of Teachers of Mathematics in the United States over the same age range. They emphasized the purpose of Statistics was to 'organize and describe information' with graphs, tables and descriptive statistics. With charts, graphs and tables they divided the questioning into four areas:- (i) Comparisons (ii) Direct Reading (iii) Interpolation & Extrapolation (iv) Problem Solving. They found that performance in areas (i) & (ii) tended to improve with age and that there were some subjects in the lower age ranges who were highly competent in these stages. Subjects tended to acquire the facility to handle items in (iii) & (iv) as they approached the later ages but a substantial proportion of even the oldest failed to cope with these more complex skills. No questions were asked which involved subjects constructing their own graphs. Questions on descriptive statistics were only presented to subjects in the older age groups. They report that less than half were able to calculate the central measures, though questions which used 'untechnical' terms were more successfully completed(e.g. 'middle' for median). Some testing was carried out on the appropiateness of measures (e.g. A shoe shop can only stock one size - which of the mean, median or mode should it stock?) and the response to these questions was poor. Questions on problem solving involved weighted means, the affect of additional data etc. The results from these were poor indicating low problem solving ability, but apparently subjects were not perturbed by answers which commonsense indicated were wrong.

1.4 Work on Probability

Since the early work of Piaget much has been written on the development of concepts in probability. More recently extensive work in this area has been carried out as part of the Probability Concepts Project undertaken by Green [1982]. Although examining areas outside of that proposed in this study the Project nevertheless provides a useful model on which to base a survey. The main purpose of Green's research was to survey the probability intuitions of children and to establish patterns of development. The actual test used was developed over two years with six pilot versions and the final test was administered to 4000 subjects of whom 2390 were finally used in a stratified sample based on General Reasoning Ability tests. Tests were read out to younger pupils and to those who might have difficulty in reading. Pupils were then allocated to one of three 'concept levels' according to how far they had progressed, questions having been designated previously as indications of certain levels of achievement. The results were also compared with Piagetian stages of development.

Although the specific probability findings are not relevant to this project, some useful conclusions can be made from this work. Firstly Green notes that development is more dependent on General Reasoning Ability than other factors such as age or sex, though boys did perform significantly better than girls. He also notes, in line with similar findings in Science, that most pupils fail to reach the Piagetian formal operations stage. Green comments that diagrams seem to help greatly with understanding and that pupils' abilities to express probability ideas verbally are weak. This tends to support the idea that many of the difficulties encountered in such tests are of understanding the question and knowing what is required. Green also notes that with more complex problems often a range of strategies were used to bring about the correct solution.

1.5 Small Scale Investigations

A number of much smaller scale investigations have been carried out into pupils' understanding of statistics. These have been generally carried out on small groups of'students' who are undertaking courses connected with statistics and although the surveys are limited in their general application they do provide a deeper insight into areas only partially examined in the larger surveys, and attempt to look at understanding rather than learnt techniques.

Work has been carried in various science research projects and others in Mathematics on interpretation of diagrams including those giving statistical information. Thornton[1986] in a recent unpublished pilot study in this field establishes four levels of understanding of statistical 'graphics':- (i) No ability (ii) Simple plotting & interpretation (iii) Interpolation, more difficult scales and more than one source (iv) Continuity, accurate interpretation & criticism. Questions were allocated to levels in the manner of Green and the test was applied to a limited sample. Results indicate that the levels were fairly well defined and that pupils at different ages did not vary greatly in the level thay had achieved. Although it is not intended in this research to investigate diagrams in depth, in view of the extensive work carried out elsewhere, this nevertheless provides a useful model for research.

A study by Barr[1986] asked technician students to calculate measures of location from a frequency table. Although there was an element of misunderstanding over the information shown in a frequency table, many subjects still made errors in calculating the measures. Again Barr points out that this was partly due to only a 'surface understanding of the vocabulary'. Other areas of difficulty arose in calculation such as failing to order unordered data for the median, or giving the 'middle' item/group irrespective of frequencies.

Jolliffe[1986] comments, though without empirical evidence, that students have difficulty in reading frequency tables when groups are of uneven size. She also notes the confusion caused by comparing percentages when totals are different, for example the failure to recognize that a 5% proportion from a total of 20 is less significant than a 5% proportion from a total of 100. This is clearly a lack of understanding of the basic idea of ratio.

Goodchild [1988] has more recently carried out detailed investigations by interviewing a small number of 13/14 year olds regarding their understanding of the mean. When asked what was understood by the words 'average contents 35' on a box of matches, students generally gave answers indicating that the mean was a rough measure of location. Few used the term 'expected' but when asked to bet on the number of matches in a particular box few chose 35. Pollatsek et al[1981] also found similar responses in their work. Further questioning asked pupils to identify the distribution of 100 boxes. The younger pupils showed no pattern in their responses but older pupils tended to opt for a bell-shaped distribution. He also comments that students did not seem to realise that a larger sample meant more proportionate variation.

1.6 Analyses of Mental Processes

Pollatsek et al[1981] have come to the conclusion, from their research, that most students have a computational knowledge of means but that this is 'minimal instrumental understanding'; they have little idea when it is appropriate to calculate the mean or when the mean gives a reasonable answer. They have used questions involving weighted means, where phrasing of the questions is such that the calculation required is not necessarily obvious. Predictable mistakes are made when calculating weighted means. They have broken down knowledge of a term such as the mean into three categories:-

(i) Functional - a meaningful real world concept i.e. what does the mean tell an individual about the data. Pollatsek et al found that this type of knowledge was generally poor, though when dealing with more concrete quantities such as weights subjects performed better at this level.

(ii) Computational - knowledge and ability of a computational formula. Individuals have a feeling for the numbers and an understanding of the basic number rules.

(iii) Analog- visual or 'kinesthetic' images, for example seeing the mean as a balancing instrument. Whilst not necessarily enabling a numerical solution to be found, this knowledge would prevent a conceptual mistake from being made.

Mevarech[1983] extended the work of Pollatsek in looking for the reasons why some of the 'misconceptions' noted had occurred. He felt that some of the difficulties with weighted means problems might result from a failure to realise that the normal rules of number do not apply to means, for example they cannot be simply added. His research justifies this conclusion and the lack of group operability of means & variances was a major source of problems. Many students in fact realised that they did not have the correct answer but did not know how to compute the correct solution.

Loosen et al[1985] have carried out some interesting research into students' intuitions of the standard deviation (S.D.) The S.D. is a measure of spread which relates all the data to a central measure i.e. the mean. It was felt that many students did not realise this fact and that textbooks did not emphasize this enough. Students were given data in the form of groups of blocks of varying length and asked to indicate which were more spread out. Two fundamental mistakes were made

(i) that the S.D. depended on the size of the blocks

(ii)data well-grouped away from the mean would have a smaller S.D.than data less well grouped, but around the mean. This indicates that the S.D. is thought of as an absolute rather than relative measure of dispersion.

Lovie & Lovie [1976] have looked at the ability of their students to estimate means and variances. This they claim is fundamental to analysing their understanding of statistics as it is unfettered by 'statistical routines which merely follow a learnt algorithm and may even confuse'. Students were examined by computer timed testing. Their results indicate that the estimates are influenced by such factors as sample size, the values of the other parameter and the size of the parameter. For example when a set of data had a higher mean than another set it was automatically assumed that the set had a higher S.D. Familiarity with the task however improved performance. It is not clear from their work how familiar the students were with the standard deviation which would clearly affect their judgement. This study indicates however that being able to estimate parameters shows some intuition as to the basic function and purpose of the two measures in terms of centrality and spread i.e. Pollatsek's analog stage.

1.7 Work on Bivariate Data

One area on which there has been scattered comment or research is in the comprehension of bivariate data techniques. This is surprising since the National Curriculum [Appendix 1] requires this to be introduced at early secondary stage. Kahneman et al [1982] have brought together many articles on the psychology of 'prediction'. These investigations considered mainly people's intuitive ideas about correlation in everyday situations rather than detecting correlation from hard data or applying particular techniques. They used previous research to maintain that laypeople have a poor ability to interpret 2 x 2 tables, looking at totals rather than proportions. Katz[1985] maintains, although unsupported by empirical evidence, that simple 2 x 2 contingency tables are easily interpreted in terms of association at an elementary level. Perhaps what is even more elementary is the interpretation of scatter diagrams with respect to correlation. Garfield & Ahlgren[1987] list some misconceptions they have discovered over time among which they include students' tendency to overestimate correlation if it is expected and vice versa, and the classic fallacy of implying cause and effect. In the light of comments by Lovie & Lovie, in the work in the last section on estimating means & variances, this appears to indicate that individuals have an intuitive expectation about the value of a measure before actual calculation which may or may not be correct.

Mosteller et al[1981] have investigated students' ability to draw regression lines through data. Students were given a transparency with a line on it and asked to place it over scattergrams with different regression lines and degrees of correlation. They found that students tended to minimize horizontal differences and that accuracy deteriorated with lower correlation. They also noted that some students consistently gave a line too steep in relation to the least squares line, whilst the intercept was well estimated. Again it is not clear whether this 'natural' skill has been interfered with by teaching or whether it is purely intuitive.

1.8 Other Topic Areas

The other topics to which reference has been made are the problems of obtaining 'unbiased' samples and expected areas under normal curves. Whilst the concept of unbiasedness is perhaps more closely related to the concepts of probability, clearly the idea of a sampling distribution of a statistic should be considered in the understanding of the mean and is a fundamental stepping stone to more sophisticated ideas such as sampling.

Rubin et al [1990] identify two apparently contradictory principles in statistical inference:-

(i)'sample representativeness' where a sample has identical properties to the population

(ii) 'sample variability' where all samples vary from the population and therefore never have the same characteristics.

They found that most students new to statistics tend to favour the 'representativeness' notion, though true understanding of inference requires a balanced measure of both principles.

Whether it is intuitive to have some notion of what proportions lie in the various regions of a Normal curve is less clear. Though familiarlity with graphs from data handling experience should make the shape familiar, there are other concepts such as using area under curves to calculate proportions which confound the issue.

1.9 Conclusions

These investigations show a rather fragmented attempt to examine various elements in Statistics. Although they provide some useful insights into the difficulties experienced by some pupils, only in certain areas (e.g. Green's work) are there attempts to propose a structure for how thinking develops in a young person. Lovie & Lovie's comment [1976] that

> "Intuitive statistics is ....concerned with the behaviour of a human being when he is required to appreciate and make inferences about an environment which is presented to him in numerical form"

sums up what we are really trying to examine in looking at the understanding of statistical concepts. In relation to this there are however some fundamental questions to be asked. For example, at what point do pupils understand the difference between spread and central tendency? It is pointless trying to investigate whether a pupil can calculate a standard deviation if the concept of spread is not understood. The ability of students to estimate the S.D. with practice noted by Lovie & Lovie might merely be a response to reward with no in depth understanding of what the figures stand for. In order to test these basic concepts it will be necessary to design a procedure which overcomes language influences and examines intuitive ideas of what the various elements of statistics represent, rather than the ability to carry out routine algorithms to arrive at an answer.

The various extracts begin to establish the structure of statistical ideas. The different areas examined show that there are several components which, while they have common elements, form separate branches within the overall area of Statistics. In addition, the various difficulties highlighted by these researchers gave details which were useful in establishing a structure for the thinking process.

The methodology used in these earlier researches was critically examined with reference to how it might usefully be applied in this project. In some cases specific questioning has been repeated where this was felt appropiate to this project.

Chapter 2

The Conceptual Development of Children's Thinking

2.1 Introduction

In order to establish a framework of concepts a detailed review was carried out on past research on the underlying developmental psychology of children, both generally and with particular reference to mathematics. The work can be categorised into three related areas:-

(i) general theories on the nature of a concept,

(ii) theory pertaining to the general development of thought processes in children,

(iii) theory relating to establishing the specific thought processes followed in the development of subject related concepts.

2.2 Psychological Background

The psychology of a child's mental growth is a vast field and it is not considered appropriate here to discuss the breadth of literature published in this area. A brief review of current views in this area was however felt necessary as a precedent to more specific work in the area of mathematics.

Mick & Brazier [1980] summarize many of the theories proposed in recent years under what they call 'constructivist' theories. Constructivism states that:

" ...... conceptual structures are spontaneously and gradually constructed through the learner's active participation in abstraction and generalization processes. The resulting conceptual structures then reorganize themselves through reflective processes"

According to Mick & Brazier all such theories originate from the work of Piaget & Inhelder and they conclude that all children pass through three distinct stages of mental development.

(i) Where cognitive structures have their origins in actions. These are of two types (a) individual actions and (b) co-ordinated actions. These merge into one another setting up a correspondence between one action and another and form the basis of all logical structures.

(ii) Where the actions are internalised as mental representation or images. This is called the 'pre-operational' stage. Transformation of reality then occurs by means of internalised actions that are grouped into coherent and reversible systems. The second level within this stage, called 'concrete operational', is distinct in that it deals with operations instead of images or static states. Concrete operational thought is characterised by "an extension of the actual in the direction of the potential". However thought still remains attached to empirical reality.

(iii) Where there is a reversal of the direction of thinking between reality and possibility. In this stage of 'formal operational' thinking, thought begins with the formation of hypotheses which are then later empirically verified. These systems result from both structural integration and combinatorial systems. Structural integration occurs where previously unrelated schemes are consolidated to form new concepts. The most important feature of this formal stage however is the existence of combinatorial systems whereby all possible outcomes of an experiment can be logically determined. At this level it is no longer necessary to relate thought to objects, and individuals can describe relations between relations.

Dienes[1967] claims that Piaget was the first to see that the process of forming a concept took far longer than had previously been believed. Dienes maintains that experiences are gradually built into a meaningful whole which leads to a concept 'clicking into focus' or being understood. The individual then tries to apply this new concept to new circumstances in order to test it and this in turn leads to a whole new crop of concepts. Dienes notes that there are two ways in which these experiences can be connected:-

(i) Conjunctive - where experiences are seen to occur together,

(ii) Disjunctive - where events are necessarily mutually exclusive.

In the classic Piagetian model it is assumed that these stages develop within approximate age bands. There have been attempts to prove that passage between stages is not necessarily bound by age constrictions. Bruner[1960] in particular has expounded the view that provided material can be presented in an amenable form age is no barrier, but this theory has not yet been satisfactorily supported by empirical evidence and there is considerable criticism of this viewpoint. The Cognitive Acceleration in Science Education project(CASE) as described by Adey[1988] and Adey & Shayer[1992] has attempted to show in recent years that the onset of formal operational thinking can be brought forward with appropriate teaching. Their 'intervention' lessons create 'cognitive conflict' whereby individuals are confronted with situations which are discordant with previous experience and understanding. Their evidence suggests that this does not always lead to new conceptualizations but that "without conflict there can be no accommodation". Their most recent findings (1992) suggest a bimodal effect with some individuals benefiting greatly from such intervention whilst others show no significant difference.

2.3 Intuition

Whilst some researchers accept the basic validity of the Piagetian structure others argue that it does not give the whole picture of intellectual development. Fischbein[1975] points out that

> "the intellectual development of the individual involves more than assimilation and organisation of conceptual systems. In addition to such structures, whose dynamics are explicitly determined by definitions and combination rules, intellectual activity involves cognitive and problem solving modalities which are less explicit, though not necessarily more primitive."

More specifically, 'intuition' denotes the existence of long-verified mechanisms, stabilised by experience. Concepts can be acquired through ordinary learning whereas intuition is formative and requires conviction based on a feeling of inward necessity. Fischbein subdivides intuitions into Primary Intuitions, which are derived directly from the experience of the individual, and Secondary Intuitions which are formed through the process of education. Skemp[1986] states that intuitive intelligence is difficult for an individual to explain.

> "Being able to do something is one thing; knowing how we did it is quite another."

Skemp refers to 'creative mental activity' in which an intuition enables a new idea to be produced in an unconscious and involuntary way.

This has important implications if interview testing is to be used, since a child who cannot explain how s/he arrived at a particular result does not necessarily indicate that the process was carried out with a lack of understanding.

Fischbein defines three levels of intuition which closely resemble the Piagetian stages, namely:-

(i)Pre-operational - synthesis of experience to confer speed, adaptability and efficiency on an appropriate action. These are essentially Primary Intuitions. Fischbein warns however that experience which is confirmed and fixed in intuitions can confer false validity on an erroneous interpretation or prediction.

(ii) Operational- coming to a conclusion given that a given set of axioms is true. Fischbein says that the major need of intelligence at this level is for quantifiable predictions. Children in this stage have a deterministic view of life

(iii) Post-operational- where past experience leads to a rapid solution without concrete operation. The 'stable, structural schemas which select, assimilate and store everything in the experience of the individual' provide the basis for extrapolation which, claims Fischbein, is the essential characteristic of intelligent behaviour.

Fischbein points out that once the basic cognitive schema are established at about 16/17 years of age, modifications to the intuitive substrata are difficult if not impossible. This suggests that misunderstandings at the earlier stages, if not corrected, can readily distort concepts at the higher level.

2.4 Developing Mathematical Concepts

One of the most comprehensive analyses of the nature of concepts in mathematics is by Skemp[1986]. Although his work relates specifically to mathematics many of his theories apply in the wider field of concept development. Skemp describes a concept as

> "a lasting mental change as a result of abstracting, which enables us to recognise experiences as having similarities to an already formed task."

Initially Skemp makes the distinction between habit learning (rote memorising) and learning involving understanding. This was the underlying principle applied later in his work when he considered concept development and setting out the test procedure. For example, rather than testing whether subjects have remembered a specific algorithm which enables them to calculate a mean from a set of data, they were more concerned with whether the subject understands what the answer means in terms of the data. Resnick and Ford[1981] define 'computation' as consisting of two elements:-

(i) simple associations - ie. the basic number rules

(ii) algorithms - a number of steps in a fixed sequence which achieve a desired result.

They claim that the bulk of elementary school experience, despite educational reform, is still of the computational type.

Skemp states that a new concept has been learnt when the resulting combination of previously held concepts results in properties which would have been difficult to predict from the previous concepts. A new concept also integrates existing knowledge and acts as a tool for further learning and makes understanding possible. He defines two types of concept:-

(i) Primary - derived from sensory-motor experiences eg. red, car, heavy, etc.

(ii) Secondary - those abstracted from other concepts. The concepts of a higher order, he maintains, cannot be communicated by simple definitions but only by collecting together suitable examples for the students to experience. Definitions only add precision to the boundaries of a concept. This again had important implications with regard to testing since the knowledge of a definition or its recognition, as has been used in some previous testing, does not necessarily indicate an understanding of the underlying concept.

2.5 Representation of Concepts

Skemp also points out that the naming of a concept, that is a sound or a mark which invokes the concept, might not be associated with the concept until after the concept has been learnt. This means that a student, who is unable to calculate a 'mean' when asked, but can when it is called the average or when reminded how to do so, may have acquired the concept though s/he has not yet made the association with the name. Skemp maintains that the use of symbols is essential in the development of concepts, as higher level concepts can often only be communicated in terms of the symbol or name of an earlier concept, it being impossible to relate to a particular concrete experience. One such example is the standard deviation. Although a student may understand the basic concept of spread it is difficult to reason in concrete terms the need for squaring of deviations to derive a suitable measure. Skemp also makes the distinction between visual & verbal/algebraic symbolization. Visual symbols are able to show certain abstract ideas such as shape and show more integrated structures. Verbal symbols however are more precise and can more easily communicate a specifically intended idea. Resnick & Ford [1981], quoting Bruner, say that the mental 'representation' of concepts occurs in three modes:-

(i) Enactive - where past skills are represented by motor skills e.g. tapping fingers on the chin to count,

(ii) Iconic -storing pictures of past operations e.g. remembers rearranging blocks,

(iii) Symbolic- a symbol or mark that in no way resembles the concept is used e.g. the number 8 looks nothing like eight objects.

In the work in this project no particular research has been undertaken to investigate a student's ability to represent concepts. However due note of these notions was made in testing to ensure that the level of representation did not exceed the level of the actual concept.

2.6 Difficulties in Concept Acquisition

Later, in the discussion of Cumulative Learning Theory, it is suggested that the acquisition of a concept needs to be measured in terms of task performance. Care, however, needs to be taken in the test procedure to ensure that concept learning was being judged rather than the application of a rote-learning method lacking any real understanding. Skemp points out two difficulties in 'schematic' learning, that is, a theoretical system consisting of a particular pattern of development by which all individuals develop a particular set of concepts. Firstly, in some cases, it is easier to learn a set of rules than develop a specific concept. That is, many students may learn how to calculate the mean without having developed the concept of the mean, either as a result of ineffective teaching or inabilty to grasp the concept. Secondly, in the learning process new information is often ignored if it does not fit in with existing schema. There is a strong tendency to preserve existing concepts even though evidence suggests they are no longer valid and there is a lack of adaptability in most individuals. This may explain why many people even into adult life find fractions difficult because they contradict the basic number concepts e.g. multiplication as repeated addition. When analysing difficulties in statistical ideas this had to be borne in mind e.g. the difficulties experienced by students with weighted means noted by Pollatsek[1981]. Garfield & Ahlgren [1987] point out that difficulties in understanding concepts arise because:-

(i) they are unlike anything the student has met before(i.e. they do not have the necessary prior concepts to grasp the idea)

(ii) they encounter interference with intuitive ideas they already have(i.e. apparently contradict existing concepts).

A concept therefore is an intuitively understood specific idea. Testing the acquisition of concepts can be difficult as it is necessary to ensure that the testing procedure examines the understanding of that concept and not the rote-learning of an algorithm or definition. The individual concepts form part of an overall pattern of thinking, for which this project set out to produce a structured model.

2.7 Concept Structure

Most of the theory of the structure of overall thinking in the area of mathematics revolves around classic Piagetian lines. Abele[1987] summarizes much work in this area in German institutions in terms of three stages of mathematical thinking:-

(i) Learning through experience,

(ii) Concept formation - looking for the "shared and essential characteristics of objects"

(iii) Representation - using graphic or symbolic representation for more formal methods.

Statistics lends itself particularly to this structure in that its basic function requires collecting appropriate material, looking for generalizations and finding appropriate symbolisation of these generalizations. Bruner[1960] essentially uses the same structure despite not accepting any imposed age limitations. It is necessary, however, to progress through the stages even with the most advanced students as Symbolic representation cannot occur without Iconic representation and Enactive stages having been reached. In other words, abstract representation cannot occur without first having the appropriate conceptual structure which in turn can only be derived from real world experience. This reflects Skemp's comments that higher concepts cannot be taught by definition but only by bringing together the necessary experiences. It is essentially this same type of structure that has been used by Green[1982] in his work on probability concepts.

In any analysis of concepts then this structure clearly needs to be superimposed. For example in looking at the concept of the mean various levels of understanding could be represented by:-

(i) looking at several objects e.g. balls of different sizes, understanding the need for a single representative measure balancing out extremes

(ii) being able to derive a mean by approximation or balancing a beam with weights spread out on it

(iii) being able to calculate a mean given the diameters of the balls.

It was considered in the present study essential that any testing examined understanding rather than rote learning.

2.8 Intuitive Thinking in Mathematics

Fischbein[1975] has suggested that, in order to solve more complex problems, certain intuitive abilities are required. Abele[1987] has established five principles which he says apply to the solving of all more intuitive problems in mathematics:-

(i) Principle of Associativity - solving a problem which allows several solutions or for which there are several methods of finding solutions.

(ii) Principle of Composing- combining one operation with another in order to gain an answer

(iii) Principle of Reversion - e.g. preservation of area when a rectangle is cut to form a square.

(iv) Principle of Transitivity - e.g. 5 + 3 = 8 hence 15 + 3 =18

(v) Principle of Variation - constructing easier adjacent tasks to solve more difficult ones.

If problem-solving can be regarded as approaching a task which is not entirely familiar then what we are looking at is how an individual attempts to use existing concepts to form new ones. If this is a valid procedure then monitoring the progress of individuals gives a valuable insight into the acquisition of new concepts. Although these ideas are not necessarily compatible with a hierarchical structure they can nevertheless be applied when setting up the hierarchies and the task definitions therein.

2.9 Cumulative Learning Theory

The final and most detailed analysis of concept development derives from what has come to be known as Gagné's Cumulative Learning Theory. White[1976] gives an excellent exposition on the development of this line of research. The basic principle is to take a target 'task' and break it down into its immediately pre-requisite sub-tasks. These sub-tasks are similarly broken down into further sub-tasks and so on until the basic skills are arrived at (Appendix 2). This forms what is known as the hierarchical structure of that particular concept. It is inherent in the structure that the target task cannot be learnt unless all the sub-skills have previously been learnt. Gagné's technique has been widely used to examine the structure of concepts in many fields. In mathematics particularly, where specific skills are more easily defined, there has been much interest. Some criticism has been made however of this rigid breakdown of the learning process (White[1976], Resnick & Ford[1981]) and the following guidelines to applying this process have been made:-

(i) Each skill needs to be defined in terms of a 'performance capability' or in terms of task-stimulus and desired response(Appendix 2B)

(ii) Subordinate tasks are included in or are components of the highest level task in the hierarchy. Earlier tasks may be 'incorporated' in higher tasks or dropped when superceded by more efficient procedures

(iii) A task's position higher in the structure does not necessarily mean that it will require more time and effort than the lower tasks. Earlier concepts may be more difficult and, once all the pre-requisite concepts are established, transfer to the higher level may be automatic.

(iv) Each sub-skill may form part of more than one hierarchy. Also skills from remote domains can contribute towards the structure of a hierarchy. e.g. map reading skills might aid graphical interpretation.

(v) More able pupils may skip pre-requisite tasks and by direct practice of sub-skills attain the target skill without any instruction. Studies have shown that teaching a complex task through a game can often bring about the learning of sub-skills as a by-product.

Thus the Gagné Cumulative Learning Theory may still be applicable, though too much dogmatic adherence to its rigidity should be avoided.

2.10 Modifications to the Cumulative Learning Model

Many researchers have criticised the use of purely linear hierarchies and claim that research using various validation techniques do not support these. Bergan [1980] describes the Ordering Theory model established by Airasian & Bart. In this they define three types of relationship that exist between pairs of skills:-

(i) Subordinate/superordinate skills - where one skill automatically precedes another.

(ii) Logically equivalent skills - all subjects able to carry out one task are automatically able to carry out another task and vice versa.

(iii) Logically independent skills - some subjects possess one skill but not another, whereas others possess the second skill but not the first.

Whereas Gagné-type hierarchial models lay down specific learning sequences, Ordering Theory compares individual links between pairs of skills. One criticism of the Gagné model is that perhaps a subset of the structure is sufficient in itself and that certain elements of it are superficial. Other factors outside the model might also influence the acquisition of a new skill which could not be accounted for in the hierarchy. A further criticism is that it does not allow for such cases where transfer to a new skill can happen when either one of two sub-skills are possessed but not necessarily both. Unfortunately in this study the scale of examining such ideas was considered prohibitive.

Ordering Theory then is not bound by the rigidity of a Gagné-type model. However in order to give some kind of logical progression it still seemed worthwhile to establish hierarchical trees, provided that not too much rigidity was assumed. The proper application of Ordering

Theory analysis would have involved evaluation processes built in to extended programmed learning and this would again be impossible with large numbers of subjects.

2.11 Conclusions

This study of theories on the development of concepts in children, whilst not specific to Statistics, provides the basic structure on which the research was carried out. From this discussion there appear to be three main avenues of thought:-

(i) That the subject of Statistics can be broken down into key concept areas. These need not necessarily be single strands leading to a single concept, but a structure which allows problems to be solved in that area.

(ii) The thinking process can be described using three stages of development under classic Piagetian lines. Though slightly different definitions are used in different literature these generally correspond to:-

Stage I: Concrete skills. Skills which can be described in terms of physically observable phenomena, e.g. choosing the middle item in respect of size of a set of different sized objects.

Stage II: Representative. Where skills are represented by the ability to complete simple number problems.

Stage III: Abstract. Where the acquisition of skills requires a degree of abstraction to solve a more complex problem.

(iii) The development of concepts can be described in detail using complex hierarchical structures. These consist of branched flow diagrams where each skill is precisely defined in terms of practical and mental tasks. Skills must be defined in this way so that understanding rather than intellectualisation is being tested. Although there are some limitations to this theory there are ways of establishing such structures.

Various methods are available to validate a proposed hierarchical structure and these will be discussed under research methodology. In addition to various testing difficulties, structures established might be merely coincidental, that is a particular sub-skill is always acquired chronologically before a higher skill though there is no link. Further, a hierarchical structure observed in a group of individuals might be a result of a particular learning strategy. Hopefully with a large widespread sample the variations of teaching systems can annul this effect. Nevertheless, despite its critics, the establishment of a hierarchical structure for various concepts should aid our understanding of the underlying thought processes.

Work on Ordering Theory and the various critiques of Cumulative Learning Theory suggest that a variety of linkages exist between concepts. In the analysis used in this research the techniques applied endeavoured to establish the nature of links between concepts rather than assume a purely linear model.

In order to establish hierachical structures for some statistical ideas it was necessary to establish certain base points beyond which it is not necessary to extend the hierarchy. For example, finding the mean of a set of numbers requires at some stage the ability to add numbers. It is not likely that it will be necessary to break down the ability to add numbers into its components (see Skemp -number rules) unless it is clear that not all of the underlying concepts are universally acquired. The hierarchies used in this research therefore start at a base point of pupils being able to carry out simple number operations.

Chapter 3

Research Methodology

3.1 Introduction

The main purpose of this project was to develop a model for the structure of the development of statistical thinking amongst secondary school pupils. It sought to examine whether there are any common patterns of learning and whether these can be described in terms of the structures described in Chapter 2. Three basic tasks were carried out as part of this study:-

(i) Hierarchical structures for the main statistical concepts were proposed. These consisted of Gagné-style hierarchies onto which were superimposed the general stages of learning.

(ii) The hierarchical structures were tested for validity as the normal pattern of development for the acquisition of those concepts.

(iii) A survey was carried out to examine the extent to which these concepts were understood and to link the various levels with specific age band.

3.2 Establishing Hierarchies

From the work in the previous chapter a hierarchical structure may be described as a sequence or set of sequences which indicate normal logical progression of thought processes which lead to the acquisition of a particular skill. Gagné, who originated the idea of hierarchical learning structures, suggests that these are most easily established by first identifying a key specific skill which is considered the pinnacle of the structure. Then the question is asked "What must the learner do to acquire this new element given only instructions". A set of sub-skills then can be derived and a similar analysis applied to these and subsequent skills until the 'base' line is reached. This procedure is known as Rational Task Analysis. To some extent the precision of the structure was not paramount at the initial stage as validation techniques should weed out any difficulties. White [1974] points out however that it is important to check for accuracy at this stage as spurious elements might be included. Resnick [1973] also states it is important that elements in the hierarchical structure are defined in terms of tasks which the subject is able to carry out.

Hierarchical structures are not appropriate to verbal knowledge. Other methods of hierarchical derivation have adopted this principle, i.e.

(i) Empirical Task Analysis (ETA):- This consists of setting subjects a particular task and observing their actions to complete it. Work has been done using precise measuring devices and recording reaction times and how these vary when certain other skills are present.

(ii) Protocol Analysis:- This is a more detailed form of ETA. Detailed records are taken of the steps taken to solve a particular problem using appropriate recording facilities. Subjects are often asked to think aloud and careful note is taken of any actions. This has the advantage that it does not require any researcher's preconceptions to establish the hierarchy.

These two methods provide a much more thorough system for establishing hierarchies. Neither are they inhibited by adult preconceptions of the way children think. However the complexity of the structures used in this research make the use of these techniques prohibitive, and as stated in Chapter 2 a subject may have acquired a concept and yet be unable to explain why.

3.3 Validation of Hierarchies

Much has been written on the subject of validation of hierarchies. The appropriate method clearly depends on whether we are adopting a Gagné-model hierarchy or using an Ordering Theory approach. The main techniques described to date are:-

(i) Scalogram Analysis:- This technique, developed by Guttman[1944], examines skills in a Gagné-type model. The responses of subjects to questions/skills on a hierarchical or 'dichotomous' scale are recorded in a table with '+' indicating correct answering or demonstration of a skill and '-' the failure to complete a task. This would result in a table of the form:-

	SKILL				
	1	2	3	4	5
Student A	+	+	+	+	-
Student B	+	+	+	-	-

By placing subjects in order of how far they had progressed through the structure, a clear pattern should emerge. If the hierarchy is valid then subjects possessing a higher level skill should also possess all the sub-skills. A coefficient of reproducibility which measures the proportion of results which comply with the designated structure can be calculated. It is the application of this technique to various cases that has led researchers to believe that rigid hierachies are not valid. Airasian, Madaus & Woods[1975] argue that the weakness of Scalogram Analysis is that it assumes a linear model and that the whole effect is a cumulative one. There are many other criticisms of linear models and the use of scalograms and Bergan[1980] lists several:-

(a) No measure is made of the magnitude of the link between skills. A mathematical test is suggested for gauging the strength of transfer.

(b) Only direct pre-requisites are examined and the process ignores reciprocal causal relationships on the same level of a hierarchy or unspecified external variables.

(c) There are cases where acquisition of either one of two skills allows transfer to the next level.

These are really criticisms of the initial model rather than of the validity of the technique. Where there are more than five items in a hierarchy the scalogram technique becomes difficult and this made it unsuitable for this research.

(ii) Ordering Theory Methods: - These form the basis of most other techniques. Since branched hierarchies may be used, the limitations of linear models are superceded. All that is measured is the extent to which one item is necessarily causal to another. A 2 x 2 contingency table can be set up using all students showing the proportions able to complete the two tasks in a particular link being examined, for example

		Skill A	
		Pass	Fail
Skill	Pass	35	0
B	Fail	23	45

For skill A to be subordinate to B there should theoretically be nobody able to complete skill B who was unable to perform A, though coefficents of reproducibility allow for some random variation. Various tests are available e.g. Baker & Hubert's method and Prediction Analysis where in the null hypothesis it is assumed that performance in the tasks is independent and that the ability to perform both is simply p(able to perform A) x p(able to perform B). The measure of prediction success can then be calculated.

(iii) Structural Analysis:- as with ordering theory the links between specific concept items are examined with a contingency table showing responses to items. The testing however is rather more rigorous. The basic hypothesis is that the number of test subjects who are able to complete the higher level task without first having attained the lower task is zero. With only one question/task per item link this would require no subjects in this category. In order to carry out an appropriate test allowing for errors several questions per item link need to be asked. Since for large hierarchies this is not practical two or three questions are the norm. For the number of questions used it is necessary to calculate the probability of the number of students actually observed to have obtained the higher level without the lower level under the null hypothesis that this should be zero. White & Clarke [1978] outline an appropriate procedure for doing this and outline a Test of Inclusion for examination of the hypothesis. A detailed analysis of this technique is described in Chapter 4.

<u>(iv) Latent Structure Analysis</u> - in this technique devised by Goodman (1974) all possible arrangements of order of task attainment are listed using a binary coding system. Chi-squared tests can then be used to compare observed data with that of all the possible models. This method although thorough and exhaustive would be unmanageable for detailed structures.

3.3 Procedure

All of the techniques above were considered for use in the initial stage of research to set up the basic structure for examination on a large scale. Bearing in mind practical considerations the following procedure was finally adopted:-

(i) Identification of the main concept areas. Rather than produce many hierarchical structures which lead to a single specific skill in the normal Gagné model it was decided to create a structure by grouping together similar skills within an area. This was done on the evidence which suggested that rigid hierarchies are rarely valid and the realisation that separation into distinct individual strands would not only be difficult but would require far more extensive questioning. Five skill areas were chosen, derived from observations of previous research. They were:-

(a) Sorting & Grouping Skills i.e. using ordering, tabulation (discrete & grouped), simple proportions and percentages to gain information about data.

(b) Statistical Measures (i) location (ii) dispersion, i.e. gaining information about data using various measures in these two areas.

(c) Bivariate Data Links i.e. basic ideas of correlation through observation of data, contingency tables, scattergrams and correlation coefficients.

(d) Sampling Skills i.e. the idea of making assumptions about a particular system using a sample of data, the concept of an error distribution, assumptions of normality.

(ii) Using a modified form of Rational Task Analysis a structure within each area was created. Discussion with others in the field was used to gain a better model before testing. On this hierarchical structure was superimposed the three main developmental structures described in the previous chapter. These were labelled I,II,III. Appendix 4 shows the pilot structures as used in the analysis.

(iii) Examination of the procedures of testing using a large group of students using written tests and practical tests. In all the techniques considered in this chapter there is very little mention of appropriate questions. In designing written tests great care needed to be taken that the questions actually tested the stated item. Mention has already been made of the students unable to calculate the mean when asked, but able to do so when it was referred to as the 'average'. In any case Resnick[1944] and White[1974] point out that verbal knowledge is not learnt hierarchically but is only 'intellectualized' skills. Visual material in tests was used to overcome this to some extent and numerical working was generally regarded as conceptual. However in many instances some articulation was required to ascertain whether a skill had been mechanically carried out without being intellectually understood. The use of multiple choice answers was found useful in these instances.

(iv) Using a crude form of scalogram style analysis and the results from initial tests the validity of the structure was examined. Diagrams showing the structure of the hierarchy were drawn for each test/student in the initial analysis on which pass/fail was indicated by shading in diamonds. These could then be critically examined to check for inconsistencies in the hierarchical structures.

These stages incorporated a pilot study in which it was hoped to not only test the basic structure of the model but examine the validity of questions. These two aims could possibly have had confounding effects. One difficulty that arose at this point was deciding whether problems derived from the model structure or from the failure of questions to test the structure adequately. However, if the structure was carefully composed, the major effects should be in the questions themselves. As a result of this early work the tests were redesigned as necessary and modifications to the structure made.

(v) Large scale testing using the modified tests/hierarchies. This main testing fulfilled two purposes:-

(a) Links between various items in the pilot hierarchy could be investigated. Structural Analysis using White & Clarke's method was applied for this purpose. This analysis enabled an examination of the strength of any links and their prequisite nature to be examined. In addition some linkage not previously considered emerged.

(b) Investigate patterns in learning progression with age. By carrying out tests on students of different ages results should indicate whether the development of skills is linked with age and in particular whether the superimposed levels had any validity. This also enabled syllabus requirements in such documents as the National Curriculum to be critically examined.

Chapter 4

A Critical Analysis of White & Clarke's Inclusion Analysis Technique

4.1 Introduction

The whole idea of rigid hierarchical structures to describe the learning process, Gagné's Cumulative Learning Theory, has been somewhat modified since its inception. The basic principle of a structured learning process, however, with the acquisition of concepts being dependent on the possession of certain prerequisite concepts, is still considered valid. However the comments on the different types of linkages between concepts discussed in Chapter 3 need to be borne in mind. The Inclusion Analysis technique devised by White & Clarke [1974] tests whether one concept is a necessary precursor to another concept. If there are two adjacent skills, Skill 1 and Skill 2, then for Skill 1 to be a precursor to Skill 2 no subject should have command of Skill 2 without also having Skill 1, although the converse may or may not be true. In Chapter 3 the various other linkages do not hold this property. The test will therefore indicate whether the particular link is of a prerequisite nature or some other form e.g. independent or parallel.

The analysis of the links requires at least two questions to be given on each task. In the analysis following the simplest case, with two questions per skill, is used throughout. In this experiment the large number of skills prohibited more questions.

Literature on this technique is scant and the only detailed exposition is in White & Clarke's original article. However their work derives from similar work carried out in genetic linkages (Rao[1950], Finney[1950], Fisher[1948]). It was therefore decided to carry out a detailed study of the technique in terms of the sensitivity of the test and its power under given conditions and the circumstances under which it degenerates.

4.2 Theoretical Background

The success rates for two skills which are thought to be adjacent can be represented by the 3x3 table:-

		Skill 2 (Questions correct)		
		0	1	2
	0	x_{00}	x_{01}	x_{02}
Skill 1 (questions correct)	1	x_{10}	x_{11}	x_{12}
	2	x_{20}	x_{21}	x_{22}

where x_{ij} are the expected number of respondents in each cell (f_{ij} are corresponding observed values).

There are certain assumptions necessary for this test:-

(i) Chance errors on all four questions are independent of one another. This seems a reasonable assumption provided in the multiple choice questions responses are randomized.

(ii) Were it not for chance errors subjects should answer questions both correctly or both incorrectly. This assumes that the skill is closely defined and that there are not subtle differences within the question regarding the skill. This also assumes that both questions do indeed apply to that skill.

(iii) Although independent, the probabilities of chance errors for each question are equal within a skill. Provided that a similar style of question e.g. multiple choice has been used for both questions on a skill, this should hold.

Let:-

P(subject answers Skill 1 questions correctly/has acquired skill 1)$=\emptyset_1$

P(subject answers Skill 1 questions correctly/has not acquired skill 1) $= \emptyset_1'$

$\emptyset_2$ and $\emptyset_2'$ are the corresponding probabilities for skill 2.

P_0, P_1, P_2, P_B are the proportions in the test sample with neither skill, skill 1 only, skill 2 only and both skills respectively.

Using these values the expected numbers of respondents falling in each of the nine categories can be expressed. e.g.

x_{00} = expected number in sample scoring 0 on both tests

$$= N[P_0(1-\emptyset_1')^2(1-\emptyset_2')^2 + P_1(1-\emptyset_1)^2(1-\emptyset_2')^2$$
$$+ P_2(1-\emptyset_1')^2(1-\emptyset_2)^2 + P_B(1-\emptyset_1)^2(1-\emptyset_2)^2]$$

and in particular

x_{02} = expected number in sample scoring 0 on skill 1 but 2 on skill 2

$$= N[P_0(1-\emptyset_1')^2\emptyset_2'^2 + P_1(1-\emptyset_1)^2\emptyset_2'^2$$
$$+ P_2(1-\emptyset_1')^2\emptyset_2^2 + P_B(1-\emptyset_1)^2\emptyset_2^2$$

The test required here uses the basic hypothesis that if Skill 1 is a precursor to Skill 2 then no subject should be able to complete questions in Skill 2 without also being successful at Skill 1. The hypotheses are therefore:-

H_0: $P_2 = 0$

against

H_1: $P_2 = k$ $(k>0)$

4.3 A method for testing the hypotheses

In an experiment there will be a set of observed frequencies f_{ij} corresponding to the x_{ij} in the table above. The normal Chi-squared test for contingency tables is not applicable here since we are examining the frequencies in only a part of the table. Nevertheless it is necessary to calculate the expected frequencies for each cell. If the P_{ij}'s and $\emptyset$'s were known these could be calculated. In the case of questions being multiple choice and with all options within each question being equally likely to be chosen by respondents without the necessary skill, values of $\emptyset_1'$ and $\emptyset_2'$ could be given. With open questions however this is not so. Two methods are possible for obtaining estimates of these parameters;-

(i) Using the cell frequencies.

Rao[1965] describes a method for using maximum likelihood estimators in a similar example in genetics. The solution even for his simpler case involves an iterative method which he himself states is clumsy and tedious.

(ii)Marginal Totals Method

This method is less accurate as it uses less information but yields a far more accessible solution.

The expected marginal totals are given by:-

Skill 1

2 Correct $T_{12} = N[(P_1+P_B)\emptyset_1^2 + (1-(P_1+P_B)\emptyset_1'^2]$

1 Correct $T_{11}= N[2(P_1+P_B)\emptyset_1(1-\emptyset_1)+2(1-(P_1+P_B))\emptyset_1'(1-\emptyset_1')]$

0 Correct $T_{10} = N[(P_1+P_B)(1-\emptyset_1)^2 + (1-(P_1+P_B))(1-\emptyset_1')^2]$

Skill 2

2 Correct $T_{22} = N[(P_2+P_B)\emptyset_2^2 + (1-(P_2+P_B)\emptyset_2'^2)]$

etc.

These expressions form a mixture of two binomial distributions. However the values of the parameters are not uniquely defined by the expressions above. In order to solve these equations further restrictions need to be placed on the parameters. $\emptyset_1'$ can be given a value in one of two cases:-

(i) Where questions involve 'open' answers the probability of an individual guessing an answer is zero i.e. $\emptyset_1'=0$.

(ii) In multiple choice questions $\emptyset_1' = {}^1/_n$ where n is the number of options in the question.

Since the critical cell is the f_{02} cell, taking $\emptyset_1'$ as zero makes the expected frequency for this cell as small as possible thus minimizing the Type 1 error. The other parameter which can be given a value to err on the conservative side is $\emptyset_2=1$. This is saying that all those with skill 2 invariably answer the questions at this level right.

To obtain the estimates of the parameters the method of moments can be used. The observed and expected marginal totals are equated with the above additional restrictions to obtain unique solutions.

For Skill 1

$$T_{12} = N[(P_1+P_B)\emptyset_1^2]$$

$$T_{11} = N[2(P_1+P_B)\emptyset_1(1-\emptyset_1)]$$

Leading to:-

$$\text{(i) } \emptyset_1 = 2T_{12}/(2T_{12} + T_{11})$$

$$\text{(ii) } (P_1+P_B) = (2T_{12}+T_{11})^2/4T_{12}N$$

Similarly with Skill 2

$$\text{(iii) } \emptyset_2' = T_{21}/(T_{21} + 2T_{20})$$

$$\text{(iv) } (P_2 \div P_B) = 1 - (T_{21}+2T_{20})^2/4T_{20}N$$

These estimates are valid under H_0 and any H_a. However under H_0, $P_2 = 0$. This in fact enables a unique solution to be obtained for the parameters (i.e. $P_0=1-(P_1+P_B)$). Expected values can now be calculated for each of the cells using the expressions for x_{ij} derived earlier. For example:-

$$x_{02} = P_0\emptyset_2'^2 + P_1(1-\emptyset_1)^2\emptyset_2'^2 + P_B(1-\emptyset_1)^2$$

Note: Since $P_2 = 0$, $\emptyset_1'=0$, $\emptyset_2=1$ this is rather simpler than the original expression.

Probabilities could be calculated for all nine cells and the probability of the observed distribution using the multinomial distribution. The test statistic would be the sum of the probabilities from all possble distributions showing a deviation from the expected values by more/less than the observed values. This would be tedious to examine.

Although all cells will be altered under different H_a, the x_{02} cell will always have the greatest proportional change. The P_2 term in the expression is $P_2(1-\emptyset_1')^2\emptyset_2^2$ and with the values of $\emptyset_1'=0$ and $\emptyset_2=1$ this will change by the value of P_2.

A significance test can therefore be carried out using a critical value based on a value to be exceeded by the observed frequency f_{02} for a particular significance value. This means that using the actual cell frequencies and the above expressions the probabilities of obtaining 0, 1 or less, 2 or less, etc. subjects in the f_{02} cell can be calculated. The critical value, x_c, of the maximum number of allowable subjects appearing in this cell before rejecting H_0 is then at the 5% level the highest value for which $P(f_{02}>x_c) < 0.95$.

The size and power are easy to calculate since f_{02} is $B(N,p_{02})$, where p_{02} is the probability of a subject appearing in this cell i.e. the size of the test can be calculated from:-

$$P(f_{02}<c) = {}^NC_0(1-p_{02})^N + {}^NC_1p_{02}(1-p_{02})^{N-1} + \ldots\ldots\ldots\ldots + {}^NC_cp_{02}{}^c(1-p_{02})^{N-c}$$

For the power under H_a: $P_2 > 0$ a new value of p_{o2} will have to be calculated by substituting the value of P_2 in the estimates of parameters equations and the expression for p_{02}. The cumulative values of $f_{02} < xc$ can then be calculated and subtracted from 1 to give the power. These procedures are fairly easily carried out using computer software such as MINITAB.

4.4 The Strength of the Test

The ability to distinguish between H_0 and H_a depends on the relative sizes of the terms in the expression:-

$$p_{02} = P_0(1-\emptyset_1')^2\emptyset_2'^2 + P_1(1-\emptyset_1)^2\emptyset_2'^2 + P_2(1-\emptyset_1')^2\emptyset_2^2 + P_B(1-\emptyset_1)^2\emptyset_2^2$$

If questions are well designed then $\emptyset_1$ and $\emptyset_2$ should be close to 1 and $\emptyset_1'$ and $\emptyset_2'$ should be close to zero. In this case the second term of the expression involves 'small' quantities to the power of four and will therefore be insignificant with respect to the other terms. Since $\emptyset_1'$ has been given the value 0, and $\emptyset_2$ the value 1, the third term is simply P_2. P_0 and P_B will of course be reduced in the calculations under H_a but this will only happen critically for large values of P_2. For a test then which is able to distinguish between H_0 and H_a it is necessary that:-

$$P_2 > P_0\emptyset_2'^2 + P_B(1-\emptyset_1)^2$$

For the right hand side to be true the following must hold:-

(i)The proportion of subjects with no skills must be low in proportion to those with skill 2 only and/or the probability of someone scoring on skill 2 without having that skill must be kept small.

(ii) The proportion of subjects with both skills must be low in proportion to those with just skill 2 and/or those with skill 1 should have a high probability of scoring on questions on skill 1.

In practical terms it is difficult to govern the proportions of subjects in the categories i.e. P_0, P_1,

P_2, P_B in a survey on a number of skills at very different ability levels. However a check on tables of frequencies should identify those cases for which the test is likely to be insensitive. The probabilities are however governed by the nature of questions being asked. Where the skill 2 questions involve multiple choice the probability $\emptyset_2'$ could be high. With complex or obscure questions on skill 1 the success level might be low and $\emptyset_1$ could cause problems. These problems were reduced as a result of pretesting procedures and c comparison between questions for a particular skill. Since the values of $\emptyset$ are found by comparison of the relative proportions scoring 0 and 1 in each skill, tables of results should indicate cases where problems may occur.

In order to examine the effect of these variables on the sensitivity of the test the size and Power were investigated for different values of $\emptyset_1$ and $\emptyset_2'$, using a fixed value for H_a and a constant critical value. These are not Operating Characteristic/Power functions but graphs which show how the Power and size are affected by a third variable simultaneously. This is similar to the way that both Power & size are affected by altering the sample size in more conventional tests.

Figure 4.4.1 shows the values of $1-\alpha$ and the Power of the test for different values of $\emptyset_1$ where N=50, $P_2=0.10$ and $\emptyset_2'=0.25$ (i.e. the probability of 'guessing' right with a multiple choice of 4 answers). A critical region of $f_{02}>3$ was used in each case.

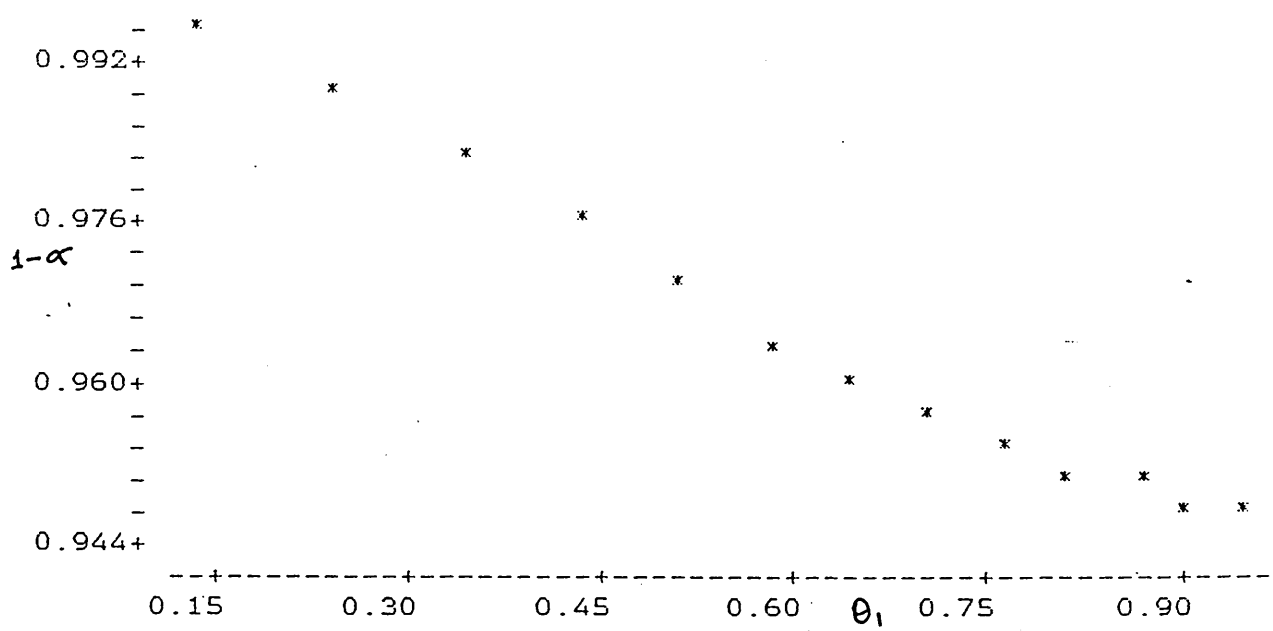
In this case it can be seen that the sensitivity of the test is altered very little by changes in $\emptyset_1$. A similar exercise was carried out using $\emptyset_2'=0.10$ and once again there was very little fluctuation in the Power of the test for a given size. In this case however the power remained unacceptable for all significance levels. This seems to indicate that the test is more dependent on changes in $\emptyset_2'$ than $\emptyset_1$. This means that in questions where difficulty was experienced and response rate was poor the sensitivity of th test is not unduly affected.

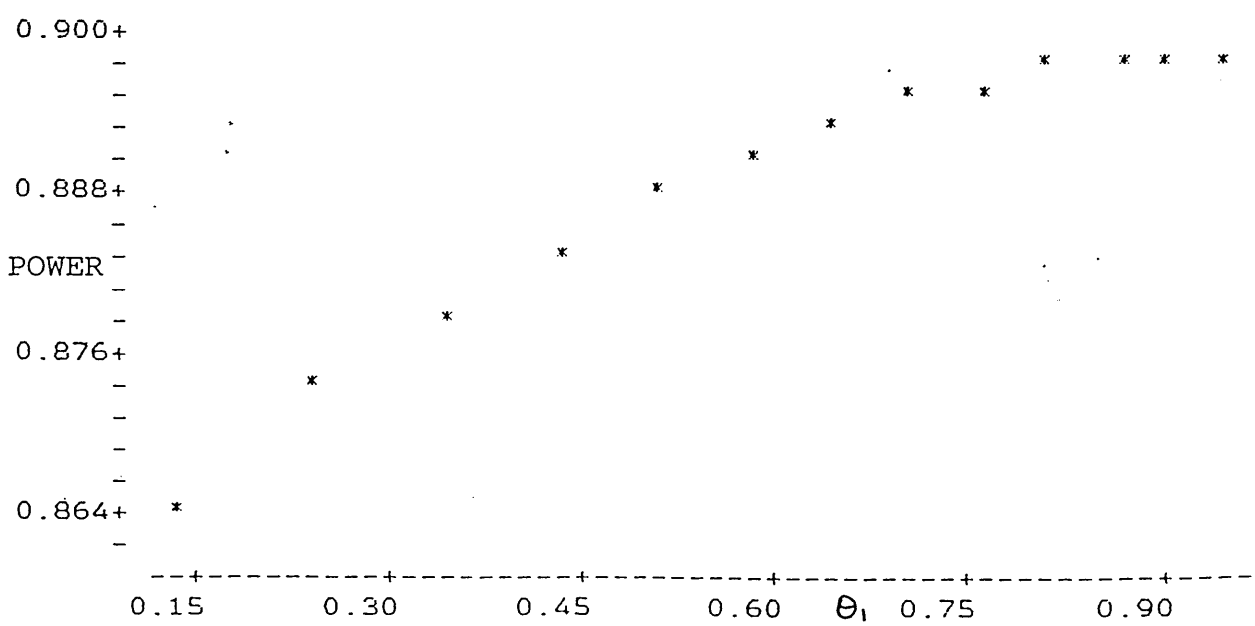
In the second series of tests, under the same conditions as above, using $\emptyset_1= 0.95$ graphs of $1-\alpha$ and Power were drawn for different values of $\emptyset_2'$. These are shown in Fig. 4.4.2. Whilst the Power is always within acceptable bounds the Type 2 error renders the test unusable with $\emptyset_2'$ greater than about 0.33. This indicates therefore that any multiple choice questions involving less than 4 possible answers, assuming all responses are equally likely to be guessed, will affect the sensitivity of the test.

In the third series of tests the effect of altering P_0 and P_B were investigated. This was rather more difficult since it was difficult to isolate P_0 and P_B as these are complex non-linear functions of the row and column totals. In the event a large number of

Characteristic & Power against $0 < \theta_1 \le 1$

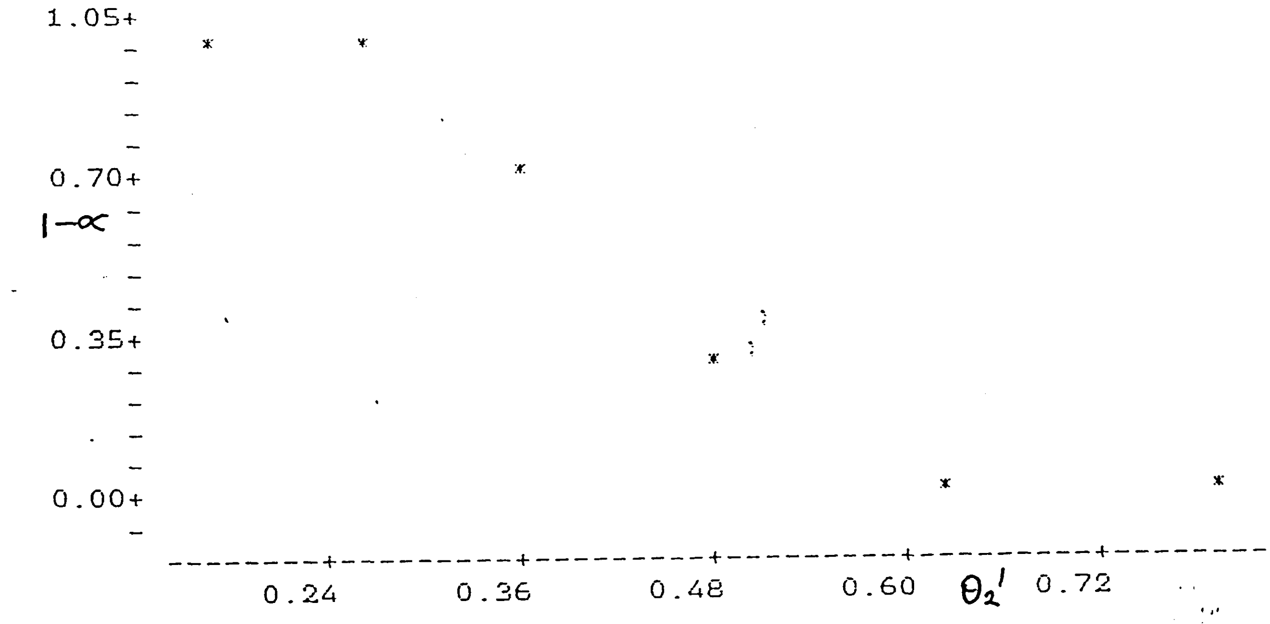
	θ_1	$1-\alpha$	POWER
1	0.945455	0.947875	0.898041
2	0.905660	0.948471	0.897801
3	0.862745	0.949460	0.897397
4	0.816327	0.950922	0.896788
5	0.765957	0.952949	0.895924
6	0.711111	0.955634	0.894735
7	0.651163	0.959075	0.893133
8	0.585366	0.963349	0.890997
9	0.512821	0.968510	0.888161
10	0.432432	0.974511	0.884393
11	0.342857	0.981161	0.879361
12	0.242424	0.987989	0.872572
13	0.129032	0.994137	0.863279

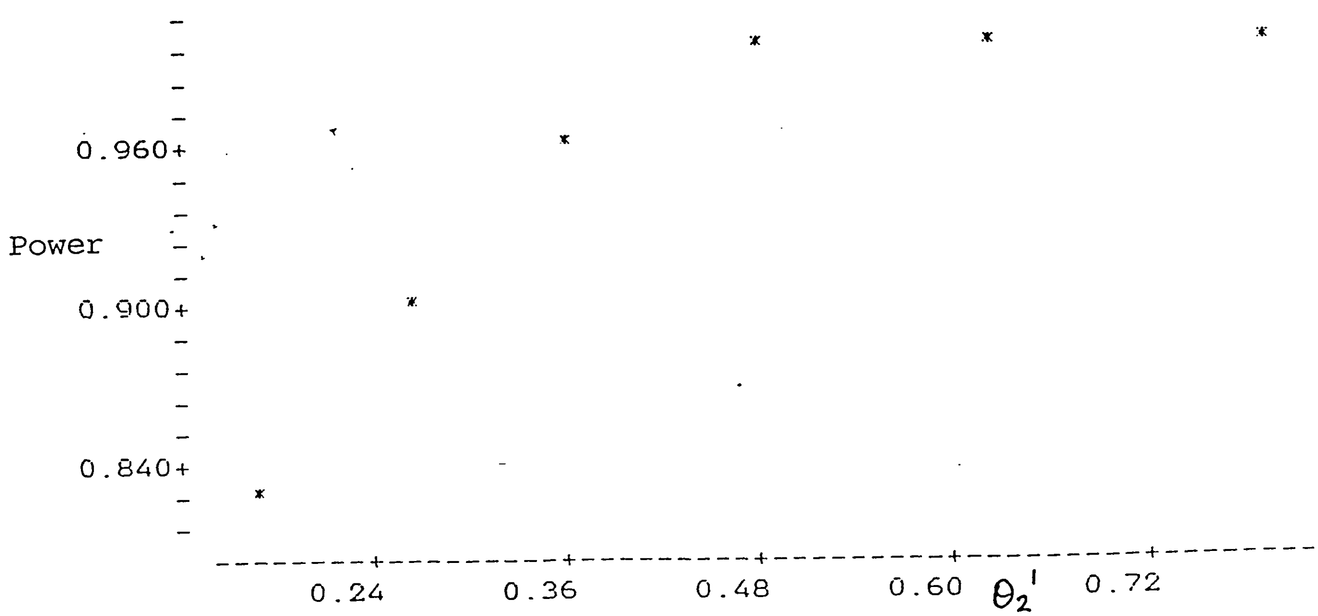




In all cases N=5 $\theta_2'=0.25$ $P_2=0.10$ C.V. Given by $f_{02}>3$

Figure 4.4.2. Table and Graph of ... Characteristic & Power against $0 < \Theta_2' < 1$





N=50 $P_2=0.10$ $\Theta_1=0.95$ C.V. given by $f_{02}>3$

Table 4.4.3. Regression Analysis of Operating Characteristic and Power

The regression equation is

$1-\alpha = 1.49 + 0.234\,\theta 1 - 1.42\,\theta 2' + 0.021\,PB - 1.61\,p0$

Predictor	Coef	Stdev	t-ratio	p
Constant	1.4916	0.2757	5.41	0.000
θ1	0.2344	0.4066	0.58	0.576
θ2'	-1.4235	0.2447	-5.82	0.000
PB	0.0209	0.1960	0.11	0.917
p0	-1.6142	0.3339	-4.83	0.000

s = 0.1426 R-sq = 82.3% R-sq(adj) = 75.9%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	4	1.03995	0.25999	12.78	0.000
Error	11	0.22373	0.02034		
Total	15	1.26367			

SOURCE	DF	SEQ SS
θ1	1	0.19614
θ2'	1	0.22898
PB	1	0.13942
p0	1	0.47542

The regression equation is

$Pw = 0.772 - 0.167\,\theta 1 + 0.486\,\theta 2' - 0.0497\,PB + 0.351\,p0$

Predictor	Coef	Stdev	t-ratio	p
Constant	0.77234	0.06373	12.12	0.000
θ1	-0.16668	0.09398	-1.77	0.104
θ2'	0.48597	0.05656	8.59	0.000
PB	-0.04967	0.04530	-1.10	0.296
p0	0.35065	0.07717	4.54	0.000

s = 0.03296 R-sq = 88.6% R-sq(adj) = 84.5%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	4	0.093188	0.023297	21.44	0.000
Error	11	0.011950	0.001086		
Total	15	0.105138			

practical examples were analysed containing a wide range of possible values. A regression analysis was carried out on both the size of the test and the Power using $\emptyset_1, \emptyset_2', P_0$ and P_B as predictors. The purpose of this analysis was to examine which variables or combination of variables, if any, had the greatest influence on the size and power of the test. The results of the analysis are shown in Table 4.4.3. The previous findings of the effect of the $\emptyset$ values are confirmed. Values of P_B do not appear to have any great effect on the sensitivity of the test. However low values of P_0 seem to adversely affect the sensitivity of the tests. Although P_0 is a complex function it clearly relates to the number in the f_{00} cell, though further investigation showed that it is dependent on the ratio between this cell and the adjacent cells f_{01} and f_{10}. The sensitivity of the test is affected when either there is a high number in the f_{00} cell and very low numbers in the two adjacent cells. Similarly when the number in the f_{00} cell is low, and the adjacent values are high, sensitivity will be affected.

On the whole these investigations show that the sensitivity is on the whole satisfactory for this test, though care must be taken in ensuring that multiple choice questions have at least four options. In cases where both skills are acquired by most individuals, or where very few possess the higher skill the sensitivity of the test is weak. In practice this usually meant the test was inconclusive in these cases.

4.5 Degenerate Cases

In carrying out the above analysis using specific cases it became evident that under certain conditions the probabilies used to find the critical cell value for f_{02} could either not be calculated or gave declining rather than increasing values.

There are two particular cases where the technique fails to give values of parameters from which satisfactory analysis can be carried out:-

(i) When $T_{11} = 0$ giving $\emptyset_1=1$

or

when $T_{21} = 0$ giving $\emptyset_2'=0$

That is, when on one or other of the skills subjects score either 0 or 2 and there are apparently no chance errors. The value of p_{02} and hence values of $f_{02}>c$ will be invariably 0 or 1 in every case respectively. This is unfortunate as this could happen where a good pair of questions were effective in their purpose.

(ii) When $T_{10} = 0$ giving a value of $P_1+P_B > 1$

or

when $T_{20} = 0$ " " $P_2+P_B > 1$

That is, when every subject gets at least 1 on a particular skill. These are less problematic in practice because if there is nobody scoring 0 on skill 1 then there must be zero frequency in the x_{02} cell and this becomes rather trivial. With no subject scoring 0 on skill 2 this suggests that there was no subject without this skill and again analysis is trivial.

4.6 Conclusion

The test appears to be a useful and easily applicable. There are certain circumstances in practice, which unfortunately may occur frequently, under which the test is insensitive or degenerate. Many of the cases can be avoided by careful question design. Multiple choice questions should have at least four response categories. Pairs of questions which yield high numbers of 1's or lack of them should also be examined carefully.

A problem that is difficult to overcome practically is where the sample contains a high number of subjects scoring two on both questions or 0 on both questions. In an ideal situation the skills should only be tested on subjects where a range of ages/skill abilities are covered so that there is a mixture of subjects with and without skills. In this survey subjects were selected within an age range where some of the early skills were universally acquired. The later skills were acquired by very few and these criteria inevitably led to cases where it was not possible to apply the test on certain linkages.

Chapter 5

Establishing Hierarchies & Initial Testing

5.1 Introduction

Having mapped out the main tasks of the research the first step was to establish some initial hierarchies which would form the basis of the study. The techniques required to do this were outlined in Chapter 4. As part of this exercise it was necessary to carry out some initial testing before a much larger scale investigation could be undertaken. The purpose of the testing was two-fold, to set up an initial model of the hierarchical structures and test the suitability of questions. The final testing could then be carried out on structures which were felt to be reasonably sound.

5.2 Drawing up Hierarchies

As stated in Chapter 3 it was decided to investigate statistical understanding under five main headings:-

(i) Sorting & Grouping Skills

(ii) Statistical Measures - Location

(iii) Statistical Measures - Distribution & Dispersion

(iv) Bivariate Data

(v) Sampling Skills

It was realised that a model with many discrete skill hierarchies was not appropriate, but as precise hierarchies were not important the use of more general skill areas was justified. It was also decided that skills would be identified under the three main areas corresponding approximately to traditional Piagetian theory, and the levels of intuition outlined by Fischbein[1975]:-

<u>(i) Stage I: Concrete Operational</u> - skills which have a direct physical result. For example, finding the median of a set of data where there is an odd number of items of data would fall into this category as the middle item can be physically chosen. When an even number of items is used however, the median no longer has an immediate physical representation. In this skill area it is important that a physical test or simple diagrammatic representation is used in questioning.

<u>(ii) Stage II: Representative/Algorithmical</u> - where a simple repesentative situation requires understanding or a simple algorithm is used to derive an answer. Examples of this are storing data in a frequency table and drawing conclusions or carrying out a simple calculation from them. Although many of the skills are

mechanical, a degree of understanding is still necessary.

(iii) Stage III: Abstract - skills involving a structured understanding and the ability to link various concepts in a judgemental way. Any mechanical skills at this level are complex and require a level of understanding to carry them out. Tasks are not just the execution of a simple algorithm, for example calculating weighted means. Interpretation involves understanding of more than one concept.

In setting up the hierarchies clear dividing lines between these levels were indicated. Even if specific hierarchical links were not rigorous it was felt that skills should be identified within the appropriate stage.

This also gave some level of comparability between separate skill areas. If, according to classic Piagetian theory, acquisition of skills within a level depended on a subject having reached a particular developmental stage, a subject who had acquired all the Stage II skills in one area should achieve a high degree of success in other areas.

Hierarchies were laid out in a 'tree' structure, as used by Gagné, with arrows linking specific skills. All skills were annotated according to the skill area and Stage -e.g. SSIIC represents a Sampling Skill at level II. The final letter was merely a cataloging mark to identify questions with no order implied. In mapping out the hierarchies, however, the order of acquisition within Stages was approximately indicated by how far down the stage it was included.

5.3 Describing Skills

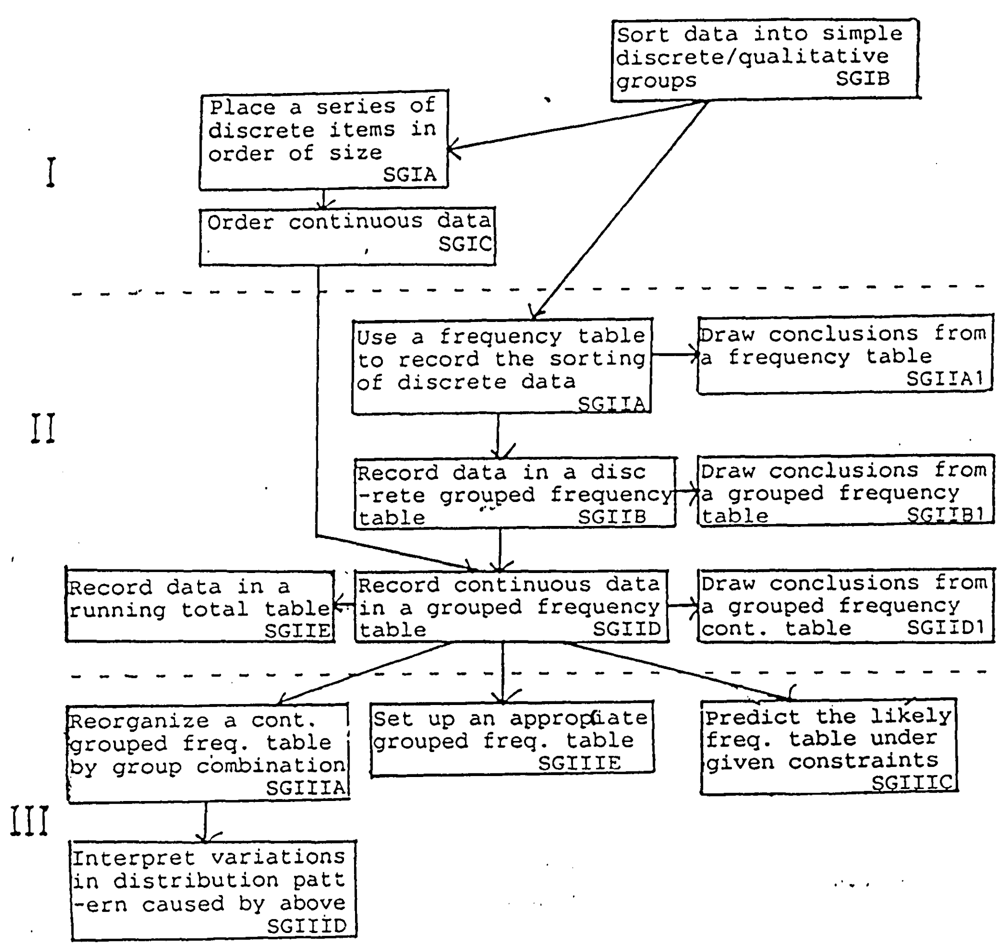
In order to place the specific skills in the hierarchies the technique of Rational Task Analysis was applied. It was difficult to ascertain specific target skills as in some areas there was more than one resulting skill. Also it was felt that an over thorough breakdown was not necessary as there were many basic skills that could safely be assumed, for example in finding means it was assumed that basic arithmetic skills were present.

Skills were defined as specific abilities. Where the level of understanding was being tested, mainly at Stage III, questions examined whether the result was expected or doubt existed as to the appropriateness.

Some skills were felt to be so similar as to be indistinguishable. In particular handling qualitative data was felt to be synonymous with handling discrete numerical data. This view was supported by questions in the initial testing stage.

Figures 5.3.1 (i),(ii),(iii) show the five hierarchies which were used prior to initial testing. These were however much adapted as a result of this testing and the expert analysis.

Sorting & Grouping Skills



Statistical Measures (i) Location

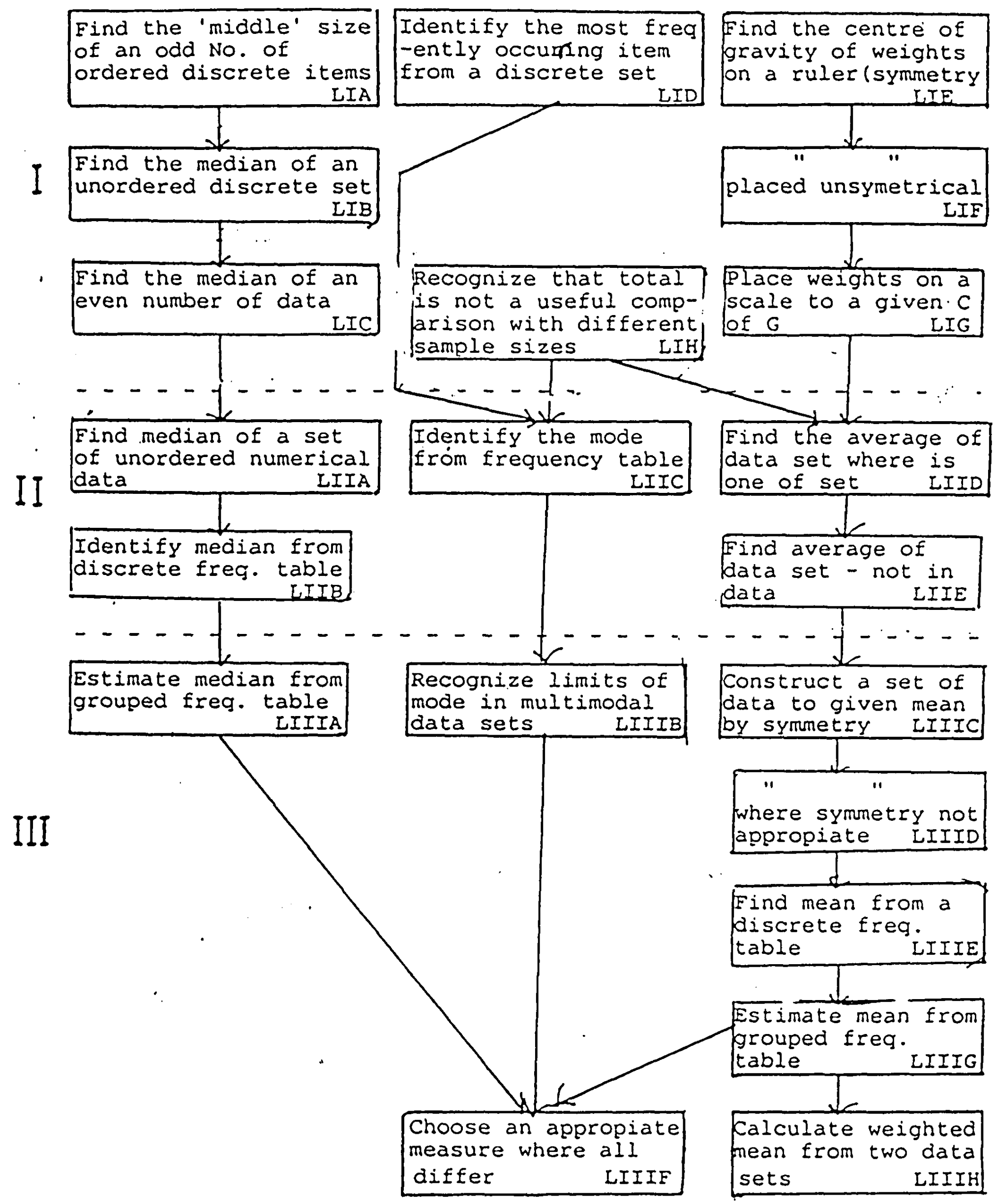


Table 5.3.1(ii) Hierarchies used in Initial Testing

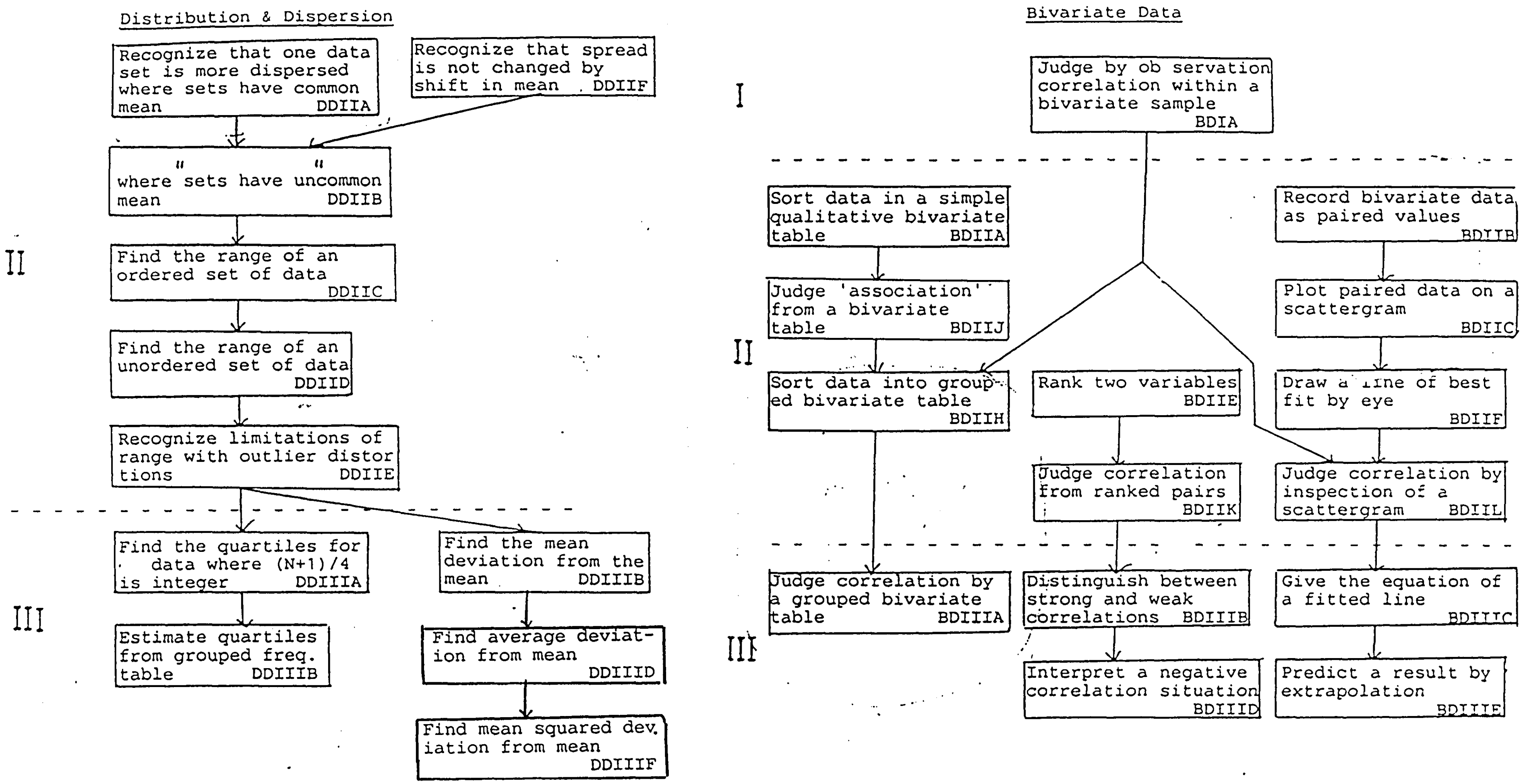
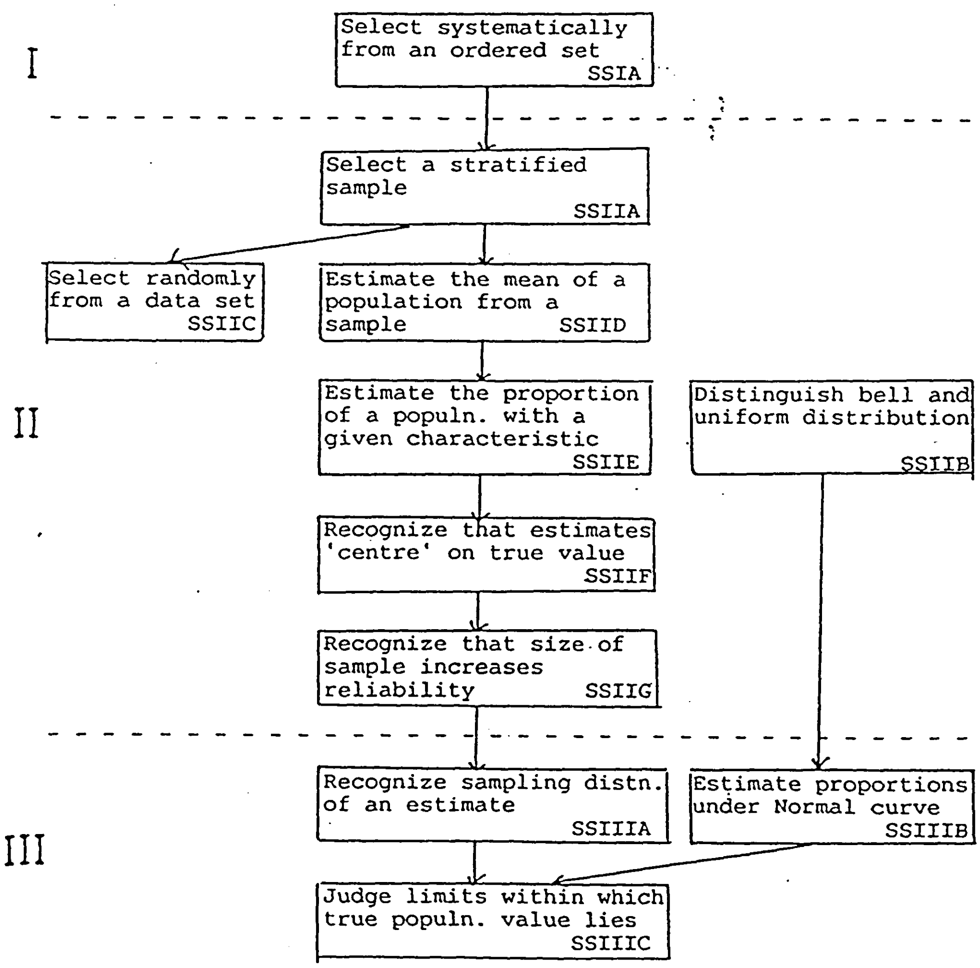


Table 5.3.1(iii) Hierarchies used in Initial Testing

Sampling Skills



5.4 Assigning Questions & Design of Tests

Allocating questions to the various skills proved to be a difficult task. The requirement of two questions per skill in order to apply the White & Clark method meant that 100 questions needed to be set. It was also felt necessary to have at least two questions per item in the pilot study to see whether the questions actually tested the same skill.

Jolliffe[1986] emphasizes the need for real data to be used, though clearly with the scale of testing proposed this was not practical. However in the Stage I questions particularly it was felt these should be based on illustrated realistic sets of data. This involved creating a file of data sets. These data sets were designed to be easily used, providing data with clearly defined characteristics. The sets were compiled using original line drawings and photocopied material. In some cases the data were usable in more than one question though data were largely compiled for specific questions. The data sets were also designed so that items of data could be selected in a variety of ways. These data sets were pretested with pupils in a classroom situation to examine the clarity and ease of use. These subjects were subsequently not used in either the pilot tests or the main testing.

Language was carefully used. In particular specific technical terms were avoided and the more colloquial term used if appropriate, e.g. using 'average' instead of 'mean'. This was as a result of research by National Council of Teachers of Mathematics in the United States[1989] which found that language played an important part in the level of understanding. In some cases however, such as using the term 'standard deviation', it was impossible to modify the language.

All the final tests were subjected to a Havering Index reading test. This involved selecting 100 words from the middle of the test, ignoring pronouns, abbreviations and names. The words were then scored according to length and a total score derived. Using a chart, which applied a linear transformation according to the number of sentences from which the words had been chosen, the score was converted to give the reading age. The results of this analysis are included in Appendix 3. The highest reading age recorded was 10.4. Consultation with Special Needs advisers revealed that any child with a reading age more than two years under his/her chronological age was regarded as in need of special help. It was therefore concluded that for those subjects aged 12 upwards the language should present no problem, although in the youngest of those studied a sizeable number may experience difficulty. In these cases the questions were read to subjects.

The tests were compiled in the form of a booklet with spaces for answers clearly indicated. This ensured that all working was included in the booklet and made the checking of scripts easier. In earlier versions of the tests there were some problems with layout, i.e. spacing of questions and data sets being presented on different pages to questions. Copies of the original tests are not included in this thesis for lack of space, but the tests used in the main testing are included as Appendix 7. These do not differ substantially from those used in the Pilot Testing. Each booklet had a frontispiece with a coded hierarchical tree on which the responses to questions could be indicated. The subject's date of birth and age were requested, but not the name.

5.5 Pilot Tests

The initial testing was carried out over a three week period in January/February 1991. All subjects were from the researcher's own school, a 12-18 comprehensive drawing from a wide catchment area and covering the full ability range. For each of the five tests 30 subjects were chosen. Classes were selected and the 5 tests were each given to some pupils in these classes. Setting arrangements made it possible to give the tests to pupils over the full range of ability within each age group. All pupils had followed for at least one year the SMP 11-16 course involving contact with statistical work.

Because different tests were administered at the same time it was possible for subjects to take the tests in a normal class situation. The tests were given out in a systematic way to ensure that no collaboration was possible and to avoid bias arising from pupils of similar ability and attitude sitting in clusters.

Pupils were told that the purpose of the test was not to test their individual ability but the overall understanding of people of their age and that this was why no name was recorded. They were given a nominal hour to complete each test and it was made clear that this was not a rigid limit. On completion of the test pupils were asked to write down the time it took to complete the test to judge the length of testing for future purposes. They were also invited to make comments on the format of the test in a blank space at the end.

5.6 Pilot Test Results

Individual performances were recorded on a coded hierarchy. These were then sorted into three age groups (i)12/13 yr olds (ii)14/15 yr olds (iii) 16+ yr olds, giving approximately ten tests in each age range.

Different problems arose in each test and thus results were initially analysed separately. Clear patterns emerged in some cases and in view of the small number in each age group, rigorous testing of all item links as proposed was felt inappropriate. Specific links were treated to more detailed scrutiny.

In analysing the results several questions were asked:-

(i) Were skills approximately in the right hierarchical position, in particular were they at the right Stage?

(ii) Were the questions appropriate to the skill and did the results show that the required skills were being tested?

(iii) Were there any problems with the test procedure which might interfere with successful large scale testing?

•

(i)Sorting & Grouping Skills

All the 28 subjects allocated this test successfully completed the majority of Stage I and II tasks. However only one in the 12/13 group and none in the 14/15 group achieved any measure of success in Stage III. The majority of 16+ subjects, however, were able to·complete all Stage III tasks. The basic structure of the hierarchy thus appeared sound and in general the ordering seemed correct.

The only 'rogue' element was SGIIE (record data in a running total table) which was only completed by those subjects achieving Stage III. This needs to be a Stage III item as it is clearly more difficult than was previously thought.

Skills SGIIA1 and SGIIB1 (interpreting frequency tables) were clearly synonymous with the appropriate main skill, i.e. subjects who are able to put information in frequency tables are able to draw conclusions from them.

Some subjects were able to reorganize grouped frequency tables given new group limits but not when asked to set up their own groups. After discussion with pupils it appeared that when regrouping tables subjects tended to work from original data rather than existing tables. This caused some question reorganization.

Subjects could generally predict a frequency table in a given situation but were not able to describe appropriate distributions. SGIIIC was felt however to be more concerned with Sampling Skills than to be listed under this heading and this item was subsequently withdrawn from the Sampling and Grouping skills schema.

As a result of this analysis a number of amendments to the Sorting & Grouping Skills hierarchy were made.

(ii) Measures of Location

The results from this area were more complex and revealed a number of problems as well as some surprising observations.

In the two younger age groups few subjects were able to achieve success in items above Stage I and even these were not completed by all subjects. In the oldest group the majority were able to perform to Stage II. There were clearly three strands in this hierarchy leading to the three main measures of location and in the proposed hierarchy these were not performed to equal standards.

In the strand leading to the calculation of the median problems arose fairly early on. Two distinct erronous strategies emerged in not only this strand but in both other strands, namely:-

(a) Using a result which was not in the data set. That is subjects found a middle item when there was an even number of data. This emerged in the LIC questions and later when subjects gave a range rather than a precise value.

(b) Where data were grouped, taking the middle group irrespective of the frequencies in each group. Although some reservations are held over the use of wording such as 'find the middle size' this does not account for all the difficulties. In view of the success in Sorting & Grouping Skills it seemed unlikely that this was due to a lack of understanding of frequency tables.

Some subjects did not order data before looking for the middle. Again, given the success rate of ordering found in Sorting & Grouping, this was surprising.

In the strand leading to the extraction of the mode most subjects achieved full success. Even amongst those only progressing up to Stage I there was a high level of success with this item. The only difficulty was that some subjects quoted the highest frequency rather than the value to which this alluded.

In the strand leading to the mean the balancing items proved more difficult than had been imagined. When symmetry (i.e. questions involved pictures of balances with weights symmetrically placed) was used subjects scored well but in unsymmetrical cases subjects found more difficulty in finding the balancing point. Only one subject who did not complete the balancing items (did not attempt them!) went on to calculate means. This was again suprising as it was thought that subjects might have a mechanical approach that gave the right answer but was not truly understood. Some younger subjects were able to calculate the mean provided the answer was within the data set or a 'real' answer. Subjects found the notion of an average of 2.4 children in a family difficult. Clearly again this reluctance to accept a non-real answer requires a high degree of abstraction and is only possible where a higher level of skill is present.

Adding data to give a stated mean was occasionally 'guessed' correctly by what was clearly a trial and error method. This item was completed in all cases, irrespective of symmetry, only by those achieving all of Stage III. Finding a mean from a discrete frequency table was sometimes completed successfully by lower level achievers and it seems that provided frequency tables are understood this skill does not require a high level of abstraction. Only the most able recognized that calculation of the mean is not affected by changing the group sizes. It was difficult to ascertain at this point whether this was due to difficulties over the grouping of data as described in the last section.

Some fairly radical restructuring of the original hierarchy was carried out as a result of the above findings.

(iii) Distribution & Dispersion

Initially it was felt that this area included no Stage I activities. It appears however that DDIIA (recognizing that two common mean data sets vary in terms of dispersion) is readily grasped by all subjects. Where means are uncommon however, subjects clearly needed tools such as the range and these tools were generally available. Skill DDIIF (recognizing that spread is not changed by shift in mean) caused some difficulty and is clearly a stepping stone to higher skills.

Strong reservations were held as to the inclusion of purely algorithmic skills such as mean deviation and standard deviation and these were subsequently withdrawn as test items. However some younger subjects asked what the quartiles were and were told that they were like the median or middle but a quarter and three quarters of the way up. This still did not enable them to complete these items and this was attributed to general problems over the concept of spread rather than mechanical completion of calculation of the quartiles.

Some care was needed with wording - some subjects interpreted 'more varied' as meaning having more distinct values rather than a wider range. However this difficulty could be due to the problems with homogeneity discussed earlier i.e. linking values to a central value rather than to each other.

(iv)Bivariate Data

There were some question design difficulties in this section. Long questions with many parts often meant that inability to do earlier tasks prevented subjects from carrying on. One question involving tadpole ages was found difficult early on and prevented further progress for many subjects.

However despite these difficulties some very useful analysis was possible. Allocation of skills to levels seemed to be on the whole appropriate and the order of acquisition proved fairly accurate. What was more apparent in this area than in any other area was the different skill levels achieved at different ages.

All subjects were able to recognize correlation in simple data sets without any formal techniques. Subjects were invariably able to complete simple bivariate tables for qualitative/discrete data. Surprisingly, in view of what was found in Sorting &

Grouping, interpreting such tables proved very difficult. An understanding of the proportionality in the tables did not appear until the middle age range (14/15) and prior to this erronous comments based on totals were common. The middle age range and above were also able to cope with grouped frequency tables and by this time interpretation skills were available and subjects had no problem in drawing conclusions from tables. Also at the middle level and above subjects were able to understand a negative correlation situation which was not comprehended by any of the younger group.

All subjects were able to rank data and from Sorting & Grouping Skills this is clearly an elementary skill. However judging correlation from ranks proved a much more difficult task. Some gave correct interpretations but clearly used raw data rather than the ranks. Questioning needed adjusting to account for this.

The ability to record data as paired values was present in subjects at all levels and putting these on a scattergram proved no more difficult. The younger children however were generally not able to draw lines of best fit and this skill seemed to be acquired at the middle age range. Once this was achieved interpolation and judging of correlation was generally understood. No subject at any level was able to give even an approximation to an equation.

(v) Sampling Skills

Again in this area there were some distinct age differences. A key task in progressing to the higher levels seems to be the ability to recognize when uniform or bell shaped distributions are appropriate in sampling distributions.

At all levels selecting systematically was not always performed well. Some interpreted 'selecting evenly' as taking the sample from only those with median characteristics. This was particularly true in the 'scouts' example (Sampling Skills Q1) where data were ordered. Selecting using a stratified system produced much better results and was coped with at all levels. Selecting randomly was difficult to analyse. Some seemed to have a used a crude stratified technique based on their observations of the distribution of data.

On the whole estimation was carried out well at all levels, though accepting a value as a reliable estimate and recognition that larger samples give a better estimate seems to be a step which makes the link between higher level skills accessible. Estimating areas under Normal curves was difficult to judge. Some subjects clearly used scales or added columns to find percentages - though it was felt that using a continuous curve would bring the additional concept of continuous distributions into play. Provided some idea of sampling distributions was understood, as well as 'reliability' of estimates, the idea of confidence limits was carried out well. It was felt however that some of the work on estimating Normal curve areas was not really testing concepts and several items were eliminated from future hierarchies.

5.7 Summary of Pilot Testing

The initial testing served its purpose very well. In general:-

(i) The form of testing and questions used were rigorously examined. There were clearly some questions which did not serve to test the skill concerned and these were either rewritten or eliminated from further testing. The general design of the test was on the whole acceptable except in some cases the test proved too long and subjects clearly got bored towards the end. The painstakingly produced visual material was well received. In comments many subjects said they enjoyed the tests and the nature of the material proved not only stimulating but accessible to all ages.

Tests on Dispersion and Sampling tended to be completed in a shorter time than other tests and this tended to encourage subjects on longer tests to give up.

(ii) Subjects' skill acquisitions were largely in line with the schema. Some adjustments needed to be made however for further testing and these fall into three categories:-

(a) Skills items in the wrong order. These were small in number and some reconstruction was necessary in all areas.

(b) Synonymous items. In some cases there were items which were clearly part of the same skill and could be combined on the hierarchy. This was useful in reducing the number of data items in future testing.

(c) Items at the wrong stage. Where there was more than one strand within a skill area stages needed to be matched up more closely. Also by comparison of the different areas some adjustment of stages was necessary.

One other noticeable element in the analysis was that in most areas there were key skills which needed to be acquired so that the intuitive leap to progress to the next level of skills could be made.

5.8 Expert Analysis

Before carrying out the necessary adjustments to the hierarchies and the questions in the tests it was decided to canvass detailed criticism of the material by a panel of experts. Five lecturers involved in Statistics Education and three secondary teachers involved in teaching Statistics at all levels were given copies of the hierarchies and the tests used. The response was excellent with detailed comments, criticisms and suggestions. The main points in the five areas are:-

(i)Sorting & Grouping: Although responses from subjects seemed to indicate that the 'draw conclusions' from frequency tables skills were synonymous with setting up the tables, there was some agreement that these were indeed separate skills and should be tested as such.

The recognition of overlapping groups was considered to be important as a precursor to the use of grouped continuous frequency tables and this was included as a skill in later tests. It was suggested that the ability to carry out 'tallying' be included as a skill though it was decided that as this was such a basic skill it was unnecessary. Initial testing indicated that no subject in fact had any difficulty with this item.

(ii) Measures of Location: Care needs to be taken with the wording in questions on the median. Asking for the 'middle' size can be misleading, and it was perhaps this that was causing subjects to opt for the middle group rather than the true median.

Some doubt was expressed as to how easy subjects would find it to mark the centre of gravity on a balance line. Ideally this should be done practically but, as the skill being tested is the ability to estimate the centre of gravity without using a trial and error method first, this was not thought necessary. In view of the findings in the initial testing it was decided to keep these balance items intact.

The effect of adding 0 to a data set was suggested as an important item in the calculation of the mean and recent US research has shown this to be an important skill. This was included as an item in later tests.

(iii) Dispersion: Questions on DDIIE (Recognition of limitations of range with outlier distortion) were heavily criticised and these were rewritten for future tests. It was suggested that a simple explanation of quartiles might enable some pupils to attempt these items, although in view of earlier comments this seemed unlikely. It was however included in further tests.

(iv) Bivariate Data: Giving the equation of a fitted straight line was considered to be too difficult and not a particularly statistical concept. This item was eliminated from further tests.

Although initial testing suggested that recording as paired values was synonymous with plotting on a scattergram, opinion seemed to indicate that these were quite distinct skills.

(v) Sampling: This section came under rather more criticism than other parts. There were strong misgivings about sampling methods. Random sampling was impossible to assess, how can one tell what method has been used? With stratified sampling it must be made clear which criteria are being used to stratify.

The ideas leading up to what is effectively a confidence limit caused concern also. Precise use of Normal curves in finding these is clearly not important. What is being looked at here is the ability to recognise the reliability of an estimated value and understand patterns of 'likely' values of an estimate.

Asking for graphs to be sketched was clearly too demanding and in future tests alternative diagrams were given and subjects asked to indicate the most appropriate one.

In addition to comments concerning particular skill areas there were a number of general criticisms/suggestions on the testing procedure.

(i) The wording of questions in some cases did not lead to the required response. In particular, questions which asked for 'reasons' were criticised as too open. This was in line with the poor response to these questions by subjects in initial tests. When redrafting the test material it was decided to replace these wherever possible by multiple choice responses.

(ii) The layout in places was too cramped with insufficient space for answers.

5.9 Implications for Further Work

In general the initial work was encouraging. Whilst some restructuring of hierarchies was shown to be necessary this was relatively minor and generally led to simplification. The revised hierarchies used for testing are included as Appendix 4.

The tests needed considerable adjustment in three areas:-

(i) rewording existing questions,

(ii) changing complete questions,

(iii) including questions on new additions to hierarchies and discarding redundant ones.

Analysing subject responses to questions involving explanations proved difficult. As a result these questions were replaced by multiple choice questions in later tests. In addition efforts were made to improve the layout of questions, giving more space for answers. In view of the shorter tests in Dispersion and Sampling, particularly after rewriting the hierarchies, it was possible, for test purposes only, to include Dispersion and Sampling in the same booklet. This not only eliminated the problem of some subjects finishing early, but enabled more tests to be carried out using the same available sample group. The tests used for the main testing are included as Appendix 7.

Chapter 6

Large Scale Testing and Analysis of Results

6.1 Experimental Procedure

The main body of testing was carried out in the months of July and September 1991. During the initial period in July those in year 11 and in the Sixth Form were involved in examinations and were not available for testing. To obtain the data for these age groups further testing was necessary at the end of September.

The researcher's own school was again used for testing. This being a 12-18 comprehensive school covering a wide ability range it was possible to take a representative sample. One advantage of using a single school was that, with a rigid syllabus in operation, learning experiences were similar and differences in performance were less likely to be attributed to the teaching process. However this may lead to patterns of understanding resulting from this specific teaching programme. The testing was carried out during the normal 70 minutes Mathematics lessons. The school sets all children in Mathematics from Year 8 within three parallel blocks in the first two years. Initial setting is done on the basis of NFER standardized scores in Mathematics administered by the Local Authority at previous schools. Minor adjustments to setting happen through years 9, 10 and 11 on the basis of a series of tests throughout each year. The sixth form has a liberal entry policy and many students return to retake GCSE at all levels in addition to those taking A level courses. There are four sets of students retaking GCSE Mathematics at Sixth Form level which are set according to the GCSE grade gained the previous Summer. At each age range sets were chosen from bands or at different examination level and it was therefore possible to gain a representative sample within each age range. In total approximately 220 tests were administered, with between 50 - 60 subjects taking each test from the full ability/age range. This meant that within each of the three age bands used (years 8/9, years 10/11 and Sixth Form) there were approximately 15 subjects. This was not a large number but seemed sufficient for analysis. None of the subjects in the large scale testing had previously been involved in the initial testing. There were few spoilt papers, the only cases were where for a number of reasons subjects were unable to complete the paper. The number of tests available for analysis in each area were:-

Sorting & Grouping	Central Measures	Distrib. & Dispersion	Bivariate Data	Sampling Skills
53	50	49	54	47

(Note: The 47 subjects who took Sampling Skills tests were those that took Distribution & Dispersion tests with two spoilt papers)

The majority of the tests were carried out under the supervision of the researcher, and in the other cases by members of staff who had been briefed in procedures. Subjects were given a brief explanation as to the purpose of the test. In particular the anonymity of the tests was stressed. Subjects were told that they might find questions which contained unfamiliar material but they should attempt these and quickly pass on to the next item if unable to complete them. Care however should be taken when questions appeared easy. No questions were to be asked unless the printing of the paper was unclear. In the event no such problems arose. Tests were allocated within the group on a systematic basis according to the groups' normal setting arrangements. This meant that no subject was sitting next to another subject taking the same test. This procedure also ensured, with natural groupings of gender and ability, that within the setted groups a representative sample of ability and gender took each test.

All subjects managed to complete the test within the given period except in one or two cases where subjects voluntarily asked to finish the test in their own time immediately after the lesson. On completion, subjects were asked to make a check on working and hand the paper to the supervisor.

6.2 Analytical Procedure

All scripts were marked personally by the researcher. This enabled any question problems, unusual answers or general difficulties to be identified. Some allowance was made for what were clearly minor numerical errors. For example when checking frequency tables compiled by subjects, particularly where measuring had taken place, some allowance was made for wrong classification. However where rather greater error had occurred this was not allowed. It subsequently emerged that subjects tended consistently to make such judgemental errors.

As in the initial tests, grids in the front of test papers were used to record performance. The grids represented the hierarchies and by shading in either half or all of the diamond, depending on whether one or two questions had been correctly answered, a good visual picture of the development of skills arose. However as more tests were carried out this became cumbersome and scores were recorded directly onto a computer file.

All scores were ultimately transferred to a computer and stored as MINITAB files. Scores were recorded in each skill as 0,1,2 depending on how many questions were answered correctly on this skill. Against each set of scores the subject's age was given in decimal form. Row numbers were marked on scripts for subsequent checking and back referencing. Using MINITAB a number of procedures were possible.

(i) Sorting data into chronological age order

This enabled data to be broken down into three age categories 12.0 -13.9 year olds (years 8/9), 14.0 -15.9 year olds (years10/11), and 16.0+. These approximately represent pre-GCSE, GCSE and post-GCSE levels, and are subsequently described as levels 0,1,2 respectively.

(ii) Searching for 'rogue'questions.

Although it was hoped that initial testing had eliminated question difficulties it was still necessary to examine the validity of questions. Skill items which contained a significant number of 1 scores were examined closely. In some cases it was decided that one of the two questions was inappropriate and results for this question were ignored. In other cases the two questions exhibited subtle differences in the skill being tested and due allowances were made.

(iii) Testing for structural problems in hierachies

Again it was hoped that initial testing had provided a reasonably accurate hierachical structure. With new positioning of skills, additional skills and a larger quantity of data a careful analysis was carried out on the structures. It appeared in many cases that groups of skills were acquired apparently simultaneously. In other cases, in terms of the analysis, all subjects had already reached a certain level and no distinction was therefore possible below this level. In order to analyse individual links, contingency tables were produced of the scores in adjacent skills:-

		SKILL 2		
		0	1	2
	0			
SKILL 1	1			
	2			

An algorithm using the marginal totals method for the White & Clarke Inclusion test (See Chapter 4) was used to find the critical number in the (0,2) cell. The general case where $\theta_2=1$ and $\theta_1'=0$ was used although in the case of the multiple choice questions θ_1' should be the probability of guessing correctly. However as some subjects gave no answer in multiple choice items this was dubious. Using a value other than 0 would also have increased the significance and for the sake of conformity over all questions the probability was kept at 0.

In view of the work in the previous chapter a careful check was made to avoid analysing pairs of skills which produced either degenerate cases or cases of insensitivity as described in Chapter 4. These were, in particular, cases where at the extremes of age and ability the majority of subjects could do neither skill or both. In some cases, though these were often degenerate, a link worked in both directions. These were cases where a few subjects scored two correct on one skill but zero on the other skill.

The linkages investigated were shown on hierarchical diagrams. Where analysis revealed that adjacent skills were not precursive a broken line was used, otherwise lines indicate data supported such linkages. Linkages with no lines show that analysis was inconclusive from the data available. Some changes could then be made to hierarchies and the data adjusted for these in later analysis. Where a skill was moved as a result of the initial analysis this has been indicated on the hierarchial diagrams in this chapter with broken line boxes showing the original position.

<u>(iv) Linking level achieved with age</u>

This was possible in two ways.

(a) The percentage success was calculated for each level I,II,III of the hierarchy. These scores could then be plotted against the ages and the degree of correlation calculated. This is slightly crude in that an individual scoring say 80% success could have done so by scoring 1 throughout or having scored 2 up to a certain skill and then subsequently 0.

(b) Looking at the percentage of subjects in each of the three age bands who had successfully attained each of the three levels. Success was defined as having achieved scores of 2 throughout, or mostly 2's but with no 1's in adjacent skills. Subjects who had scored 0 in a skill but achieved 2 in all subsequent skills were also deemed successful. In some areas it was possible to define sublevels within a level and these were similarly treated.

(v) Comparison of levels achieved across different skill areas.

Apart from Dispersion and Sampling tests no subject completed a test in more than one skill area and therefore no direct analysis of the comparability of levels in different skill areas could be made. By looking at the percentages in each age band achieving the various levels in the five skill areas, some indication as to the correspondence of the levels over the areas was possible.

6.3 Detailed Analysis by Skill Area

6.3.1 Sorting & Grouping Skills

(i) In the marking of skills SGIIA,B &D subjects were permitted to have frequencies up to +/-1 from the correct answer in two frequency categories, particularly where some form of measurement was required. Initially it was felt that perhaps this was too rigid bearing in mind that the purpose of the assessment was to see whether the subject could actually carry out an allocation process rather than produce an accurate result. Later evidence indicated, however, that subjects who were inclined to make errors in one type of table frequently had problems in allocation in the other tables. This could be due to careless working in general or a genuine difficulty in categorization.

Skill SGIIIA yielded many scores of 1. On closer examination of the scripts it appeared that the problem lay in the different nature of the questions. The first question asked subjects to re-allocate frequencies into a new table whose limits were given. The allocation could possibly have been done using the raw data but there was no evidence that this was in fact the case. In the second question (SGIIA) subjects were asked to reorganize data into a table of their own making. The latter clearly depends on being able to draw up a table with groups and later evidence suggests that this is a harder skill than at first judged.

Overall Inclusion Analysis was made difficult by the fact that all subjects were able to complete all Level I questions and most of Level II. Skill links at the two lower levels were impossible to test. It had been hoped to test whether the 'interpretative' skills SGIIA1, B1 & D1 were independent of the parent skill. Only SGIID and D1 could be tested and the evidence suggests that these skills were independent.

The skill SGIIE, recognizing non-overlapping groups, was newly introduced as an early skill in Level II. Evidence suggested that this was misplaced in this level. The overall percentage success in the three age groups for this skill and level II as a whole are shown in Table 6.3.1A

Table 6.3.1A. Percentage success of the different age bands of skill SGIIE and other skills at level II.

		Percentage success	
		SGIIE	Level II
	12.0-13.9	51%	76%
Ages	14.0-15.9	60%	88%
	16.0+	86%	97%

As a result of these findings this skill was transferred to Level III and the results adjusted accordingly. The effect of this skill on SGIID, SGIIIE and SGIIIA was further examined. SGIID proved clearly not to be dependent on SGIIE, i.e. subjects were able to put data into given groups without being able to recognize the problems of overlapping groups. However comparison with SGIIIE and SGIIIA showed that SGIIE was indeed a requirement of these skills. The difference appears to be that subjects cannot progress to creating their own frequency tables until they have grasped the concept of non-overlapping groups. In the questions on 'non-overlapping groups' subjects' most frequent error was to opt for exclusive groups i.e 6-9, 11-14, 15-19 etc. This was also evident in subjects' own attempts to create groups. Some subjects, although managing to create groups, did this very inefficiently using groups of width 1 or a single group containing all the data.

Other skills at level 3 were difficult to examine due to the poor overall success rate. Interestingly in the questions on cumulative frequency several subjects used inverse frequencies. These were often subjects at the lower ages who had not met this idea within teaching.

Boxplots were drawn of the subjects' percentage success at Levels II and III against their age.[Figure 6.3.1 (ii)] The correlation of these variables was also calculated. The plot for Level II reveals an interesting pattern. The youngest achieved success rates ranging from 25% up to 100%. As the age of the subject increases the success rate is far less likely to be low at this level. The weak correlation merely reflects that nevertheless some of the more able younger children achieved a high degree of success at this level.

At Level III the pattern is less marked but nevertheless a fairly good correlation exists between age and percentage success. In this case some older subjects had low success rates whilst some of the more able middle age range subjects achieved moderate success.

The percentage of subjects in each age range achieving full success(as defined earlier) at each Level are given in Table 6.3.1(i).

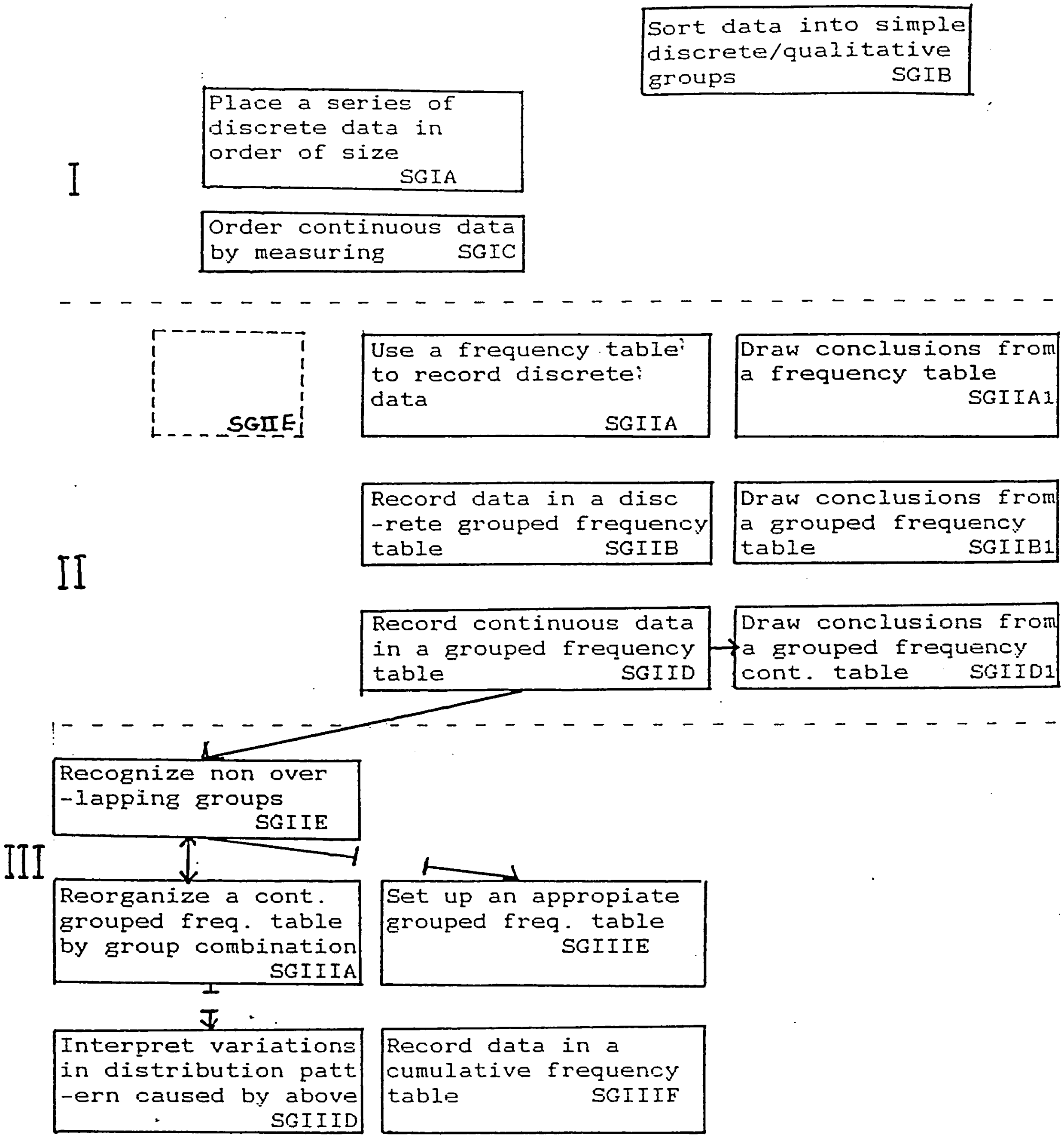
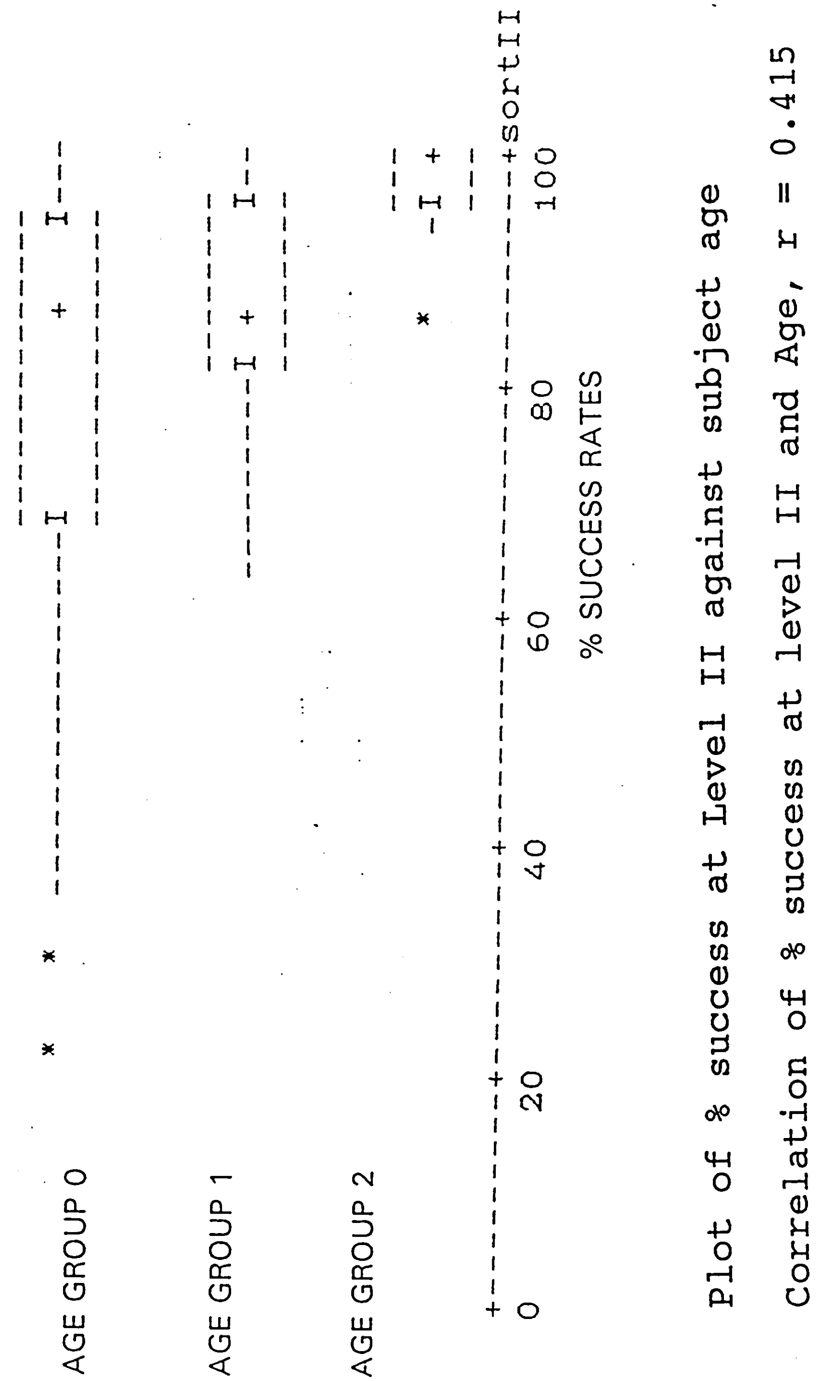


Figure 6.3.1. (i) Final hierarchical diagram for Sorting & Grouping Skills showing linkages.



0	12.0-13.9	1	14.0-15.9	2	16.0+

Figure 6.3.1. (ii) A Box and Whisker Plot of % success at Level III with respect to age group.

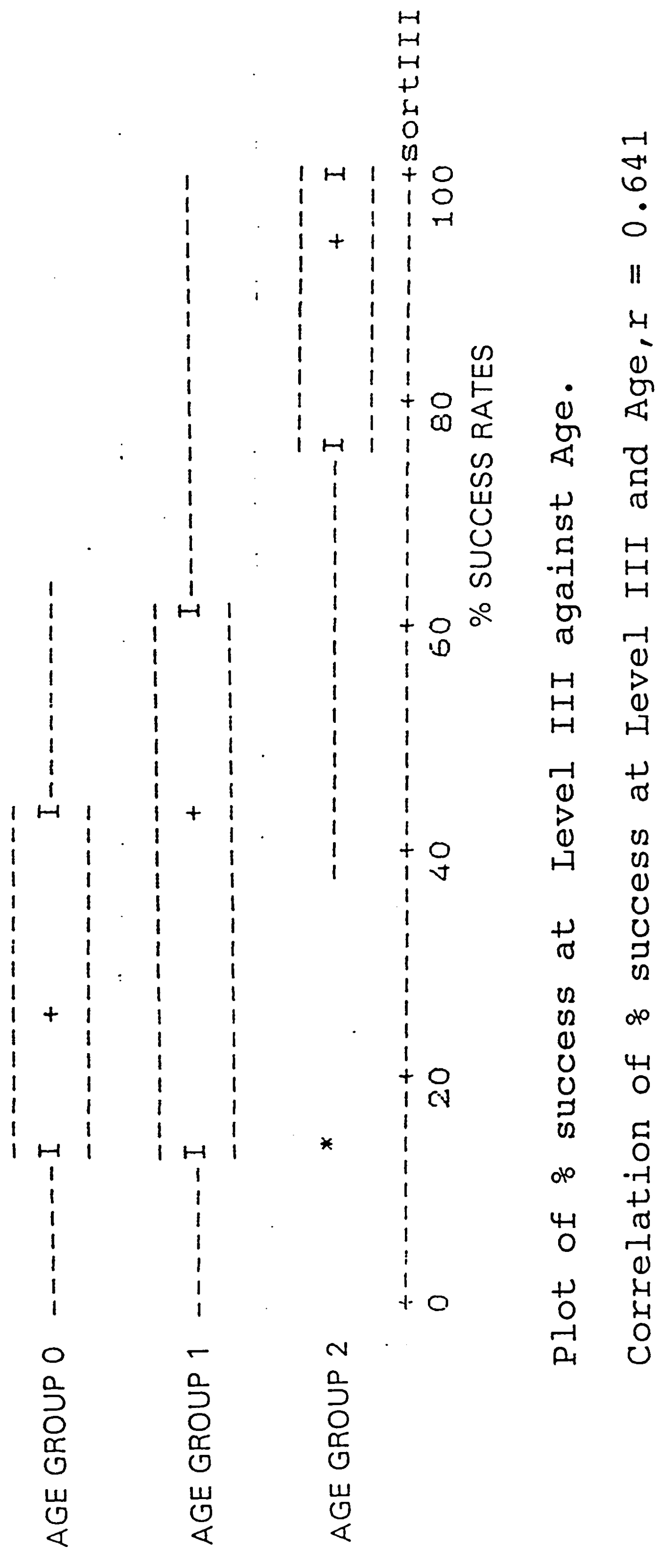


Figure 6.3.1. (ii) Box and Whisker Plot of % success at Level III with respect to age group.

Table 6.3.1(i). Maximum level achieved by subjects in each age group

	Maximum Level Achieved		
Age range	Level I	Level II	Level III
12.0-13.9	33%	67%	0%
14.0-15.9	13%	67%	20%
16.0+	0%	43%	57%

Thus none of the younger age group was able to achieve total success at Level III, whereas all those in the higher age group were able to achieve at least Level II. This further supports the evidence from the boxplots and correlation that the youngest subjects were able to reach the highest levels though the oldest were far more likely to achieve these higher levels.

6.3.2 Measures of Location

Several skills appeared from an initial analysis of the data to present problems in terms of analysis. LIG (putting weights on a beam to balance it) had a large number of 1's and close inspection revealed that the problem was in subjects being unable to deal with the unsymmetric case. That is, one question involved putting an extra weight on which could have been solved by symmetry, whereas the other needed a deeper understanding of the nature of balancing. The comparable unsymmetrical case with means was placed at Level II and clearly similar placing needed to be made for the unsymmetrical balancing case.

LIIC (identifying the mode from a frequency table) also appeared to have a large number of 1's. On investigation this showed that many subjects had given the frequency of the modal value rather than the mode itself. Whether this was just carelessness, or a genuine misunderstanding, i.e. following an algorithm looking for the highest frequency but not relating it back to the data, is not clear. The link between this and the similar situation with the median (i.e. giving (n+1)/2 rather than median) was investigated. Giving (n+1)/2 was not in evidence, but with frequency tables the middle group description was often given. Similarly with calculating means from frequency tables erroneous methods were used, group descriptions were simply averaged or descriptions correctly multiplied by frequencies and then divided. It was first thought that perhaps these errors were due to a lack of understanding of frequency tables but in view of the good performances in this area this did not seem so likely.

LIIF (finding the median of an even number of items of data) had a low response rate for its position in the hierarchy. Later evidence showed that in fact 'non-real' answers were a difficult concept. However it was felt at this point that the questions in both cases were perhaps misleading in that they implied that a specific answer was required. Even though many subjects gave an answer nowhere near the true value it was decided to eliminate this item for analysis purposes rather than promote it to a higher position because with clearer questioning this item might prove to be easier.

LIIJ (finding the mean from a discrete frequency table) similarly had a very poor response rate. However where success was evident this usually resulted in a score of 2, indicating that the poor response was because the task was more difficult than had been supposed rather than bad questioning. This skill was subsequently promoted to Level III.

Responses to LIIG (recognizing that the total is not a useful comparative measure with unequal numbers of data) showed a variety of incorrect and correct strategies.

This caused some assessment problems. Most noticeable of the strategies used to obtain a correct result were:-

(i) pairing off values

(ii) finding the average

(iii) looking at the extra values which would have to be added to one set to achieve the same total and comparing these with the other set.

Nevertheless it was felt that all the above strategies demonstrated the recognition of the problem of looking at totals.

As with other areas, the excellent scores at Level I and the very poor responses at Level III made a rigorous analysis of the hierachies difficult at these levels. However some very striking results emerged with Level II skills and some at the highest level. The two 'recognition' skills LIIG (total is not useful as a comparative measure) and LIIE (effect of inclusion of 0 in data set) were subjected to rigorous inclusion analysis to determine their position within the hierarchy. LIIG showed a very strong dependence on LIID(finding mean of data set where not in set) and similarly higher level skills with the mean were dependent on LIIG. Whether these are skills that are necessary for transference or are merely chronologically ordered in the overall development of skills of course cannot be analysed. LIIE in the same way requires a basic ability in calculating means and precedes more complex handling skills. In the questioning, however, only examples using means were used although clearly this requirement has implications in the calculation of medians and with finding modes. It would also have been useful to include questions with data sets which contained 0's which subjects had to include in calculations.

The two most prominent linkages deriving from the inclusion analysis in this skill area were between LIIH(finding the centre of gravity of unsymmetrically placed weights on a balance) and the LIID and LIIE skills(finding a mean from a set of data where the answer is/isn't in data set). The other skills using balances(with the reservations expressed above about LIG) also show a marked precedence to the tasks involving finding and interpreting means. This suggests that proficiency with a balance and weights facilitates easy access to understanding the concept of the mean.

It became evident at this point that there were certain types of skills that occurred in the branches for each type of measure and it was decided to investigate whether the acquisition of skills in these areas occurred roughly simultaneously. Two way tables and correlation were used to compare these areas. For the three skills involving frequency tables(LIIB, LIIC, and LIIJ) analysis was impossible because, apart from skills in finding the mode, few possessed any skill in this area. The same problem arose with finding answers that are not amongst the data set(LIIF, LIIIB, LIIE). This could be interpreted as showing that there are basic conceptual difficulties, such as non-real answers, which are common to all three measures. LIIG and LIIE, skills involving recognition of special circumstances, did however correspond well.

Hierarchical trees showing keylinkages are shown in Figure 6.3.2 (i).

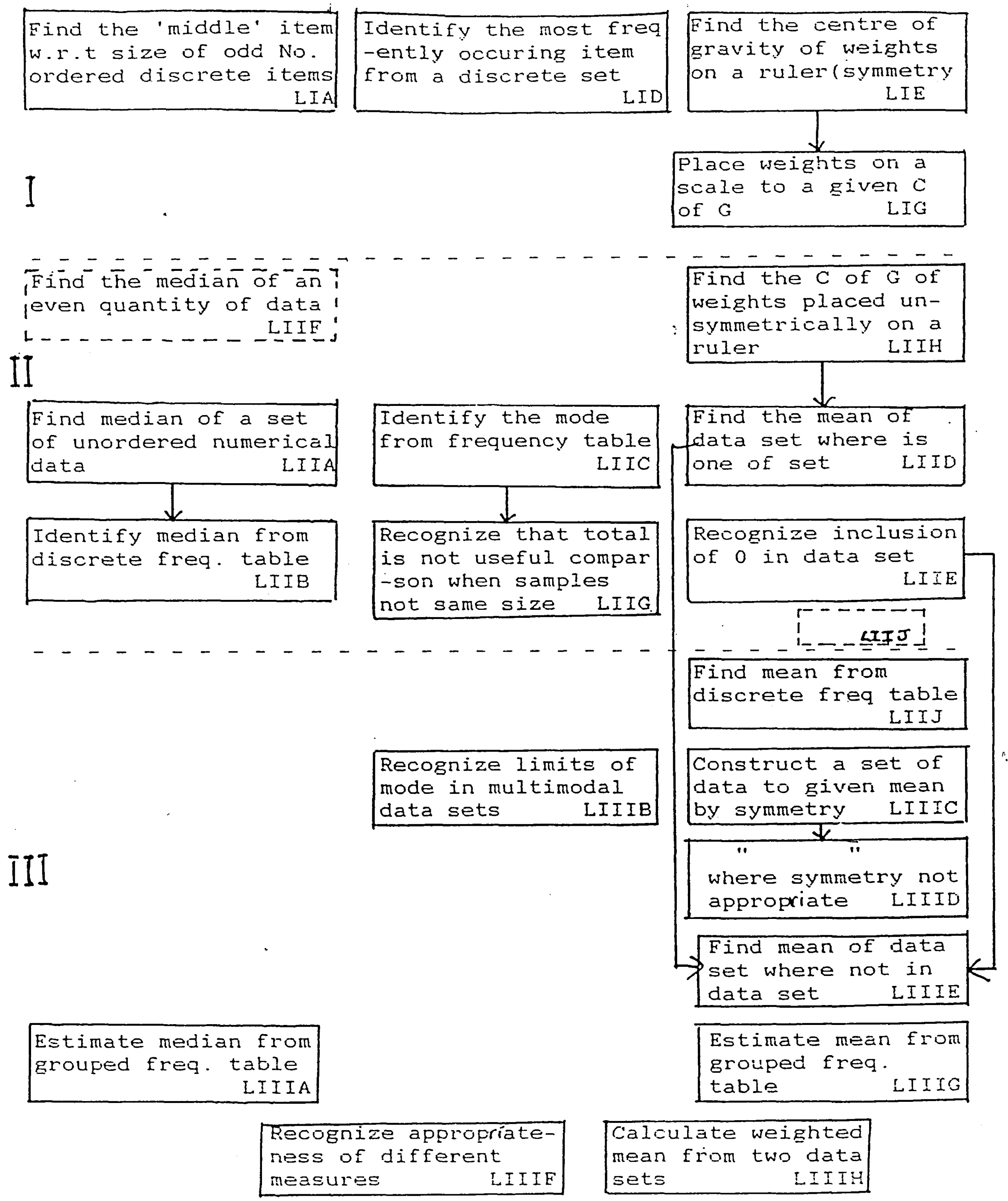
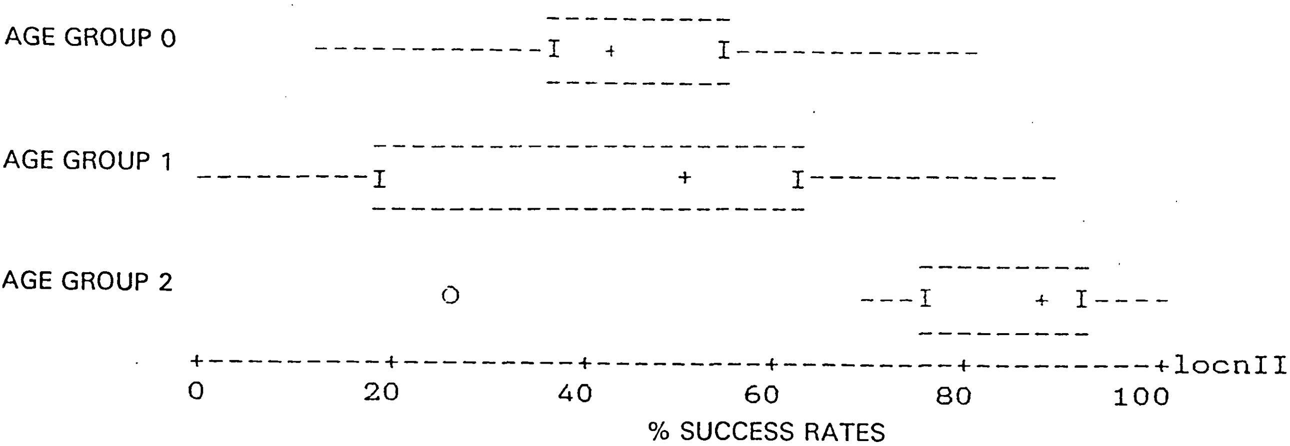
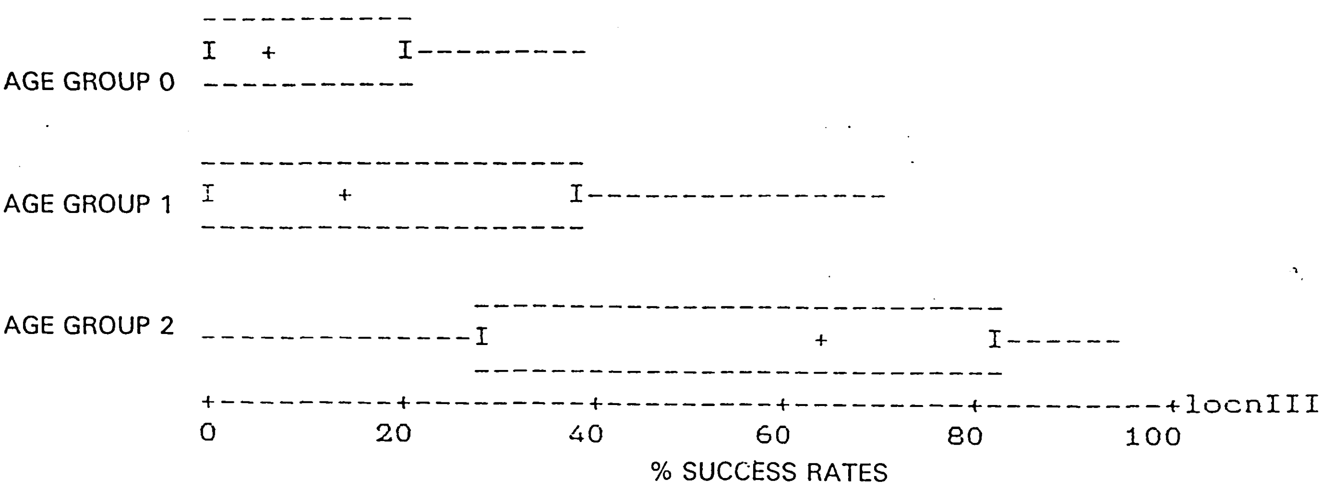


Figure 6.3.2.(i) Final hierarchy for Measures of Location skills showing linkages.



Plot of % success at Level II against Age

Correlation of % success at Level II and Age r= 0.341



Plot of % success at Level III and Age

Correlation of % success at level III and Age r = 0.427

<u>Figure 6.3.2.(ii) Correlation of ages with % success at different Levels.</u>

The development of skills by age was not as clearly evident in this area as with sorting and grouping. Plots of percentage success rates at the two levels (Figure 6.3.2. (ii)) revealed a less marked pattern and correlation was weak at Level II. With Level III a more distinct pattern emerged and correlation was higher. However with a large number of subjects up to the age of 15 gaining little success in this level some doubts exist as to the validity of the measure. Analysis of the 'success' rates at each of the three levels is given in Table 6.3.2:-

Table 6.3.2. Percentage of subjects achieving success at different levels for each age band

	Maximum Level Achieved			
AGE	0	I	II	III
12.0-13.9	20%	68%	12%	0%
14.0-15.9	18%	70%	12%	0%
16.0+	0%	38%	25%	37%

An attempt was made to link the skill areas above, i.e. dealing with 'non-real' answers, frequency tables and recognition skills, to age. But with small numbers in the older range findings were inconclusive. A similar assessment of individual skills failed to produce any age distinctions. It appears even that younger subjects achieved a slightly higher rate of success at the recognition skills.

6.3.3 Dispersion Skills

Item DDIA (recognizing that one data set is more dispersed than another) had caused some doubt earlier on as to whether this was a Level I or II skill. Although response was fairly good it corresponded well with DDIIC (finding the range of an ordered set) and as this was clearly Level II, DDIA was subsequently regarded as Level II also. Questions on finding the range proved suprisingly difficult. Only a third of the youngest subjects were able to find the ranges of ordered data sets. Perhaps a brief explanation of the term might have helped here. DDIIE (recognizing limits of range with outlier distortion) produced a large number of 1's. There was a fundamental questioning difference here in that one compared two sets of data whilst the other looked at an isolated set of data. Subjects generally responded better in the two data sets questions.

All useful links were examined by inclusion analysis. No link was found between DDIA (Recognition of dispersion differences in two sets of data) but then most subjects were able to complete both. Two strong precedences emerged, where DDIIC(range of an ordered set) appeared to be a necessary requirement of:-

(i) DDIID(range of an unordered set)

(ii)DDIIF(recognition that spread is not affected by mean shift).

The first link is not suprising; most subjects could do neither or both so it appears that for those who can find the range it makes little difference as to whether the data are ordered. Again to be able to recognize that a shift in mean does not affect the spread clearly needs a tool such as the range in order that this can be realised. The poor response to questions on quartiles made this area difficult to analyse. All key linkages are shown on Figure 6.3.3. (i)

Plots of percentage successes at Levels II and III and correlations show a similar pattern to sorting and grouping skills (Figure 6.3.3.(ii)). Younger subjects had a wide range of success at Level II but as age increased subjects were more likely to achieve a higher rate of success. With Level III many subjects at all levels had no success at all, those achieving the higher rates of success were all in the older age groups. A breakdown of those achieving each of the three levels endorses this, as is shown in Table 6.3.3A:-

Table 6.3.3.A <u>Percentage of those in each age band achieving success at each level</u>

	Maximum Level Achieved		
AGE	0/I	II	III
12.0-13.9	86%	14%	0%
14.0-15.9	83%	17%	0%
16.0+	30%	50%	20%

Some care needs to be taken in the interpretation of these figures however. Level III consisted of only two items on quartiles, a topic which most will have studied in the GCSE course (at least the more able). This probably undermines the findings in that success at level III could be as a result of having covered work on quartiles rather than having achieved a level of conceptual understanding.

One item of particular interest was DDIIE(recognition of the limits of the range with outlier distortion). This appeared not to be dependent on being able to calculate the range. The questions however were designed to test three different possible strategies for judging dispersion:-

(i) Using the absolute range irrespective of extremes.

(ii) Using the idea of homogeneity - i.e. did the results cluster around specific values (not a central one)

(iii) Relating the data to a central position.

An investigation was carried out into the likely approach used by subjects of different ages. Bearing in mind comments made earlier on the subtle differences between the two questions on this topic the responses fell into the categories shown in Table 6.3.3B:-

Table 6.3.3B <u>Percentage of subjects using particular strategies in each age group</u>

	Question 7 (Grades)			Question 8 (Xmas Box)		
	Range	Homog.	Central	Range	Homog.	Central
12.0-13.9	24	41	35	44	28	28
14.0-15.9	15	30	55	18	41	41
16.0	10	40	50	17	33	50

The findings are by no means conclusive, but it appears that older subjects are less likely to use a simple range approach and more likely to use a 'centrality' approach. This is worthy of further investigation.

Distribution & Dispersion

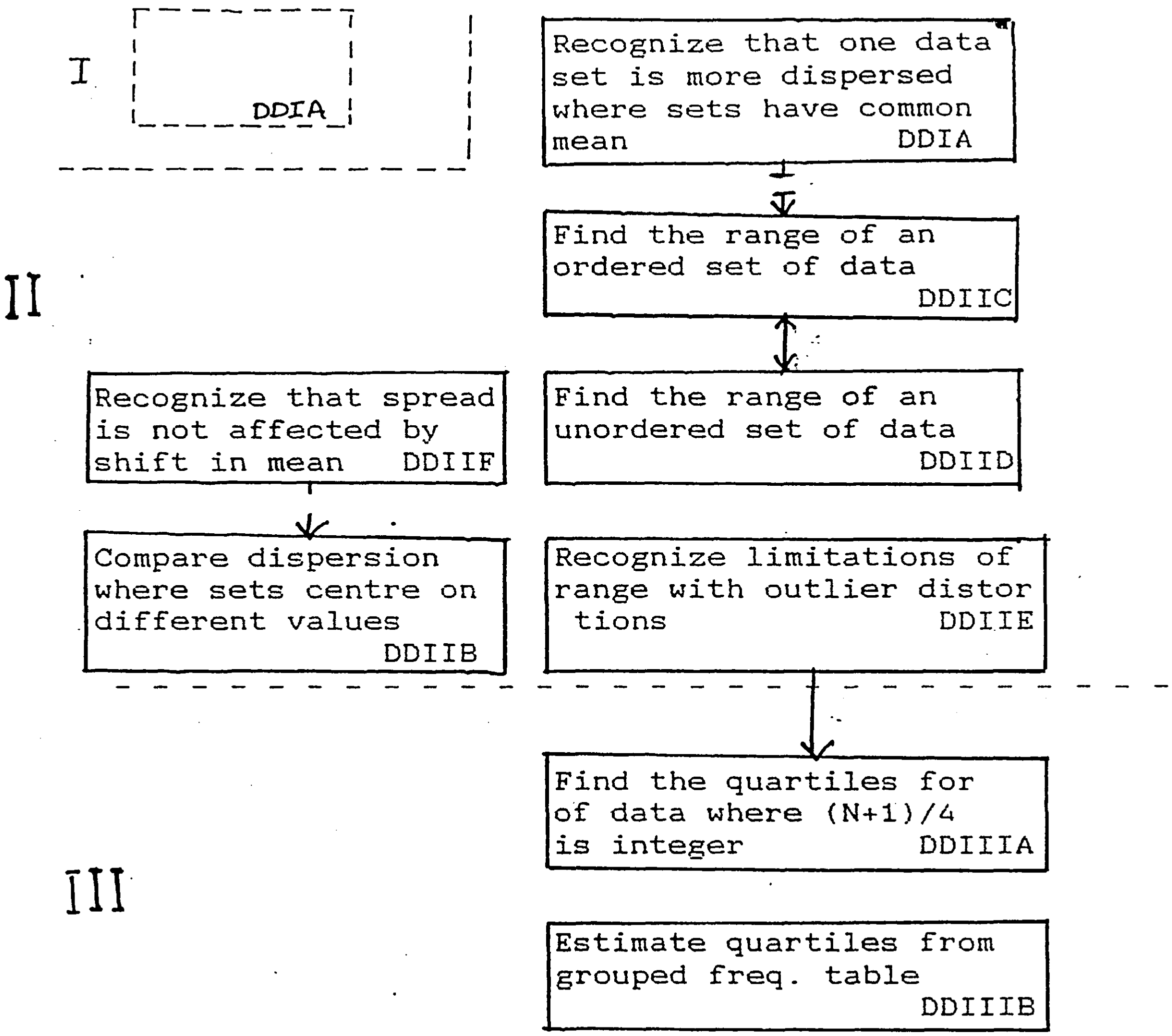


Figure 6.3.3.(1) Final hierarchy showing linkages for Distribution & Dispersion skills.

```
                            -------------------------
GROUP 0   ---------I       +                        I----------------
                    -------------------------

                              ---------------------
GROUP 1   ---------------I           +         I----------
                              ---------------------

                                                          ---------
GROUP 2                                   ------------I     + I----
                                                          ---------
          +---------+---------+---------+---------+---------+dispII
          0        20        40        60        80       100
                                   % SUCCESS RATES
```

Correlation of % success at Level II and Age, r = 0.261

```
          -
GROUP 0   +             O           O
          -

          -
GROUP 1   +             O           O
          -

GROUP 2   --------------------------
          I     +                  I-------------------------
          --------------------------
          +---------+---------+---------+---------+---------+dispIII
          0        20        40        60        80       100
                                   % SUCCESS RATES
```

Correlation of % success at Level III and Age, r =0.302

Figure 6.3.3.(ii) Plots of % success at level II and III against age groups for Dispersion & Distribution Skills

6.3.4 Bivariate Data

Despite some radical changes following the initial tests there were still some problems with the questioning and the hierarchical structure. BDIB (ranking two variables) proved more difficult than expected. It was felt that perhaps in the eggs example, because measuring was involved, subjects lost track of the question and many used actual values here whilst using ranking in the other question. However in view of the overall difficulty with this item it was moved to Level II for analysis purposes.

The long questions with tadpoles and ducks still presented problems in that subjects finding difficulty early in the question were often prevented from further progression. Without considerably lengthening the test paper it is difficult to see how this could be rectified.

Sorting data into bivariate tables proved easier than had at first been thought in both discrete cases(BDIIA) and continuous(BDIIIF) but after due consideration it was decided to leave these in their original positions in the hierarchy as they appeared to correspond well with other skills in that level. Question 4, the seatbelt problem, proved unusually difficult with not a single correct response. However the simpler case of this item, Q7 on drug effectiveness, was found easier although still more difficult than imagined. Quite why question 4 caused more problems is uncertain; perhaps a cause was the size of numbers or the larger size of table or perhaps that it involved a situation where subjects had preconceived ideas. This was treated as a rogue question and the skill moved to Level III.

Questions on skill BDIIIK (judging correlation from ranks) caused some problems as subjects who had been unable to rank often gave correct answers to this, clearly using the raw data. Correct answers to this item where no ranking was shown were ignored and the correct answers with ranking must be open to suspicion as to the method used.

Negative correlation proved little problem with over 50% success rate overall and this is clearly not a Level III skill. This was therefore moved to Level II.

Drawing a straight line through a set of data proved rather more difficult than had been expected. In the duck example (Q.5(b)) it could be argued that with so few points the line was not obvious, but even in the tadpole example few subjects perceived the straight line. This item was therefore moved to Level III. Interestingly though, in both questions, a small number were still able to give a good extrapolation value without having drawn the line. Despite these problems it was possible to reorganize the results so that useful analysis could be undertaken.

As with other areas a detailed inclusion analysis was undertaken. In some examples, where there was a high mutual success rate, the technique degenerated and either gave no solution or produced significance levels greater than 1. Similarly in other cases with a low mutual success rate seemingly very significant results were suspect. Other links such as between BDIB(ranking variables) and BDIIK(judging correlation from ranks), whilst significant were rather trivial. Among those results of importance is the ability to recognise negative correlation (BDIIID) clearly precedes (BDIIIB) 'distinguishing between strong and weak correlations'. It appears that subjects who are able to assess correlation seemed to have no more difficulty in dealing with negative correlation.

The problems subjects found in drawing a straight line on a scattergram(BDIIF) have been previously mentioned. This ability appears to be a necessary precursive skill to assessing correlation from a scattergram(BDIIL). This is confounded by the low mutual success rate in both items but nevertheless seems to suggest that the problems found in assessing correlation from a scattergram stem from the inability to percieve a line through the data. A marked order of precedence appears to exist also between BDIIF and extrapolation (BDIIIE) as expected, but again poor success rates confuse the issue here, and as stated earlier some subjects were able to extrapolate without having drawn a line. The link between recording paired values(BDIIB) and plotting on a scattergram(BDIIC) was confounded by the high mutual success rates of these two. This again suggests that there is very little difference in these activities. Valid linkages are shown in Figure 6.3.4. (i)

Boxplots for Level II percentage success rates against age (Figure 6.3.4. (ii)) reflect similar patterns to those met in Sorting & Grouping and Dispersion. At both levels a fairly good correlation exists (r= 0.436/0.425). More importantly, those in the lower ages at Level II range from low success to high success rates, whereas as age increases subjects become increasingly likely to gain greater success. A similar pattern exsists at Level III, however there is a noticable proportion of older subjects who fail to achieve any measure of success at this level. To compare the percentage at each age group achieving each level slightly broader bounds were allowed to define 'success' in view of some question difficulty. The success rates are shown in Table 6.3.4A:-

Table 6.3.4A <u>Table showing percentage success rates at each level for each age category</u>

	Maximum Level Achieved		
AGE	I	II	III
12.0-13.9	60%	32%	8%
14.0-15.9	47%	29%	24%
16.0+	33%	33%	33%

These patterns reflect the findings above.

Bivariate Data

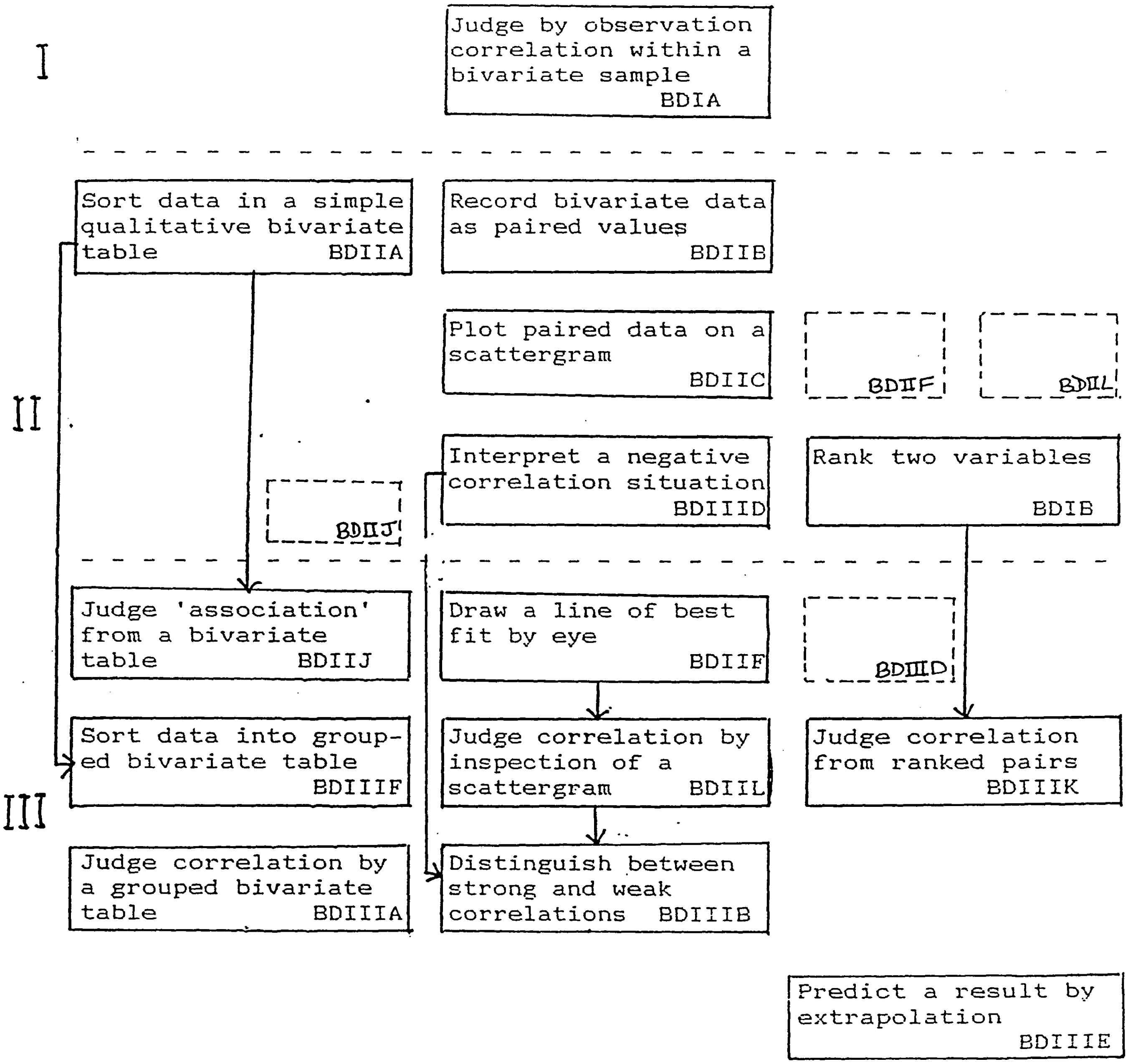


Figure 6.3.4.(i) Final hierarchy showing linkages for Bivariate Data Skills

```
                                -----------------------
AGE GROUP 0         ------------I            +        I---
                                -----------------------

                                       -------------------
AGE GROUP 1         -------------------I     +           I
                                       -------------------

                                                  ----------
AGE GROUP 2                  *               ---I   +     I
                                                  ----------
      +---------+---------+---------+---------+---------+bivaII
      0        20        40        60        80       100
                           % SUCCESS RATES
```

Correlation of % success at Level II and Age r= 0.436

```
                 ----------------------
E GROUP 0   -------I     +             I-------------------

                      -----------------------
E GROUP 1   ------------I        +            I-------------
                      -----------------------

                                ---------------
E GROUP 2   *          ---------I      +       I--------
                                ---------------
            +---------+---------+---------+---------+---------+bivaIII
            0        20        40        60        80       100
                                 % SUCCESS RATES
```

Correlation of % success at level III and Age r = 0.425

Figure 6.3.4.(ii) Plots of percentage success rates at Level II and III against Age for Bivariate Data Skills

6.3.5 Sampling Skills

The radical reorganisation of the test in this area produced a rather more consistent set of responses than in initial testing. The questions on SSIA (systematic sampling) however yielded poor responses causing doubt on its positioning in Level I. The wording of this question could have possibly caused some difficulty. Responses were far poorer in the lower age groups. It appeared that two erronous strategies were used:-

(i) picking all objects of the perceived 'median' size

(ii) using a stratified method.

An analysis of the responses showed overall the following responses:-

Median	Stratified	Systematic
33%	28%	39%

Although the questioning cannot be regarded as a pure survey on children's natural strategies for taking a sample the results do give an indication of what subjects regarded as 'fair'.

In questions on SSIIF(recognizing that estimates centre on true value) subjects often declared that their method gave a value close to the true value even though it was of a completely different order. This casts some doubt on the reasoning used in other responses to this skill. In SSIIB (recognizing sampling distributions) some older subjects perceived that a perfect distribution was unlikely and due allowances were made for this.

Inclusion analysis proved useful in this area in that it strongly supported the hierarchy in most aspects. The only notable exception is that SSIIIA(recognize sampling distribution of estimate) does not seem to require SSIIG(recognizing reliability increases with sample size). Key linkages are shown in Figure 6.3.5. (i).

The boxplot of Level II% success rates against age (Figure 6.3.5. (ii)) reveals a similar pattern to that in other skill areas and correlation is as expected. With Level III the pattern is less well defined than in other skill areas. The limited number of questions in this area and the large number in all age ranges scoring 0 could possibly account for this.

Sampling Skills .

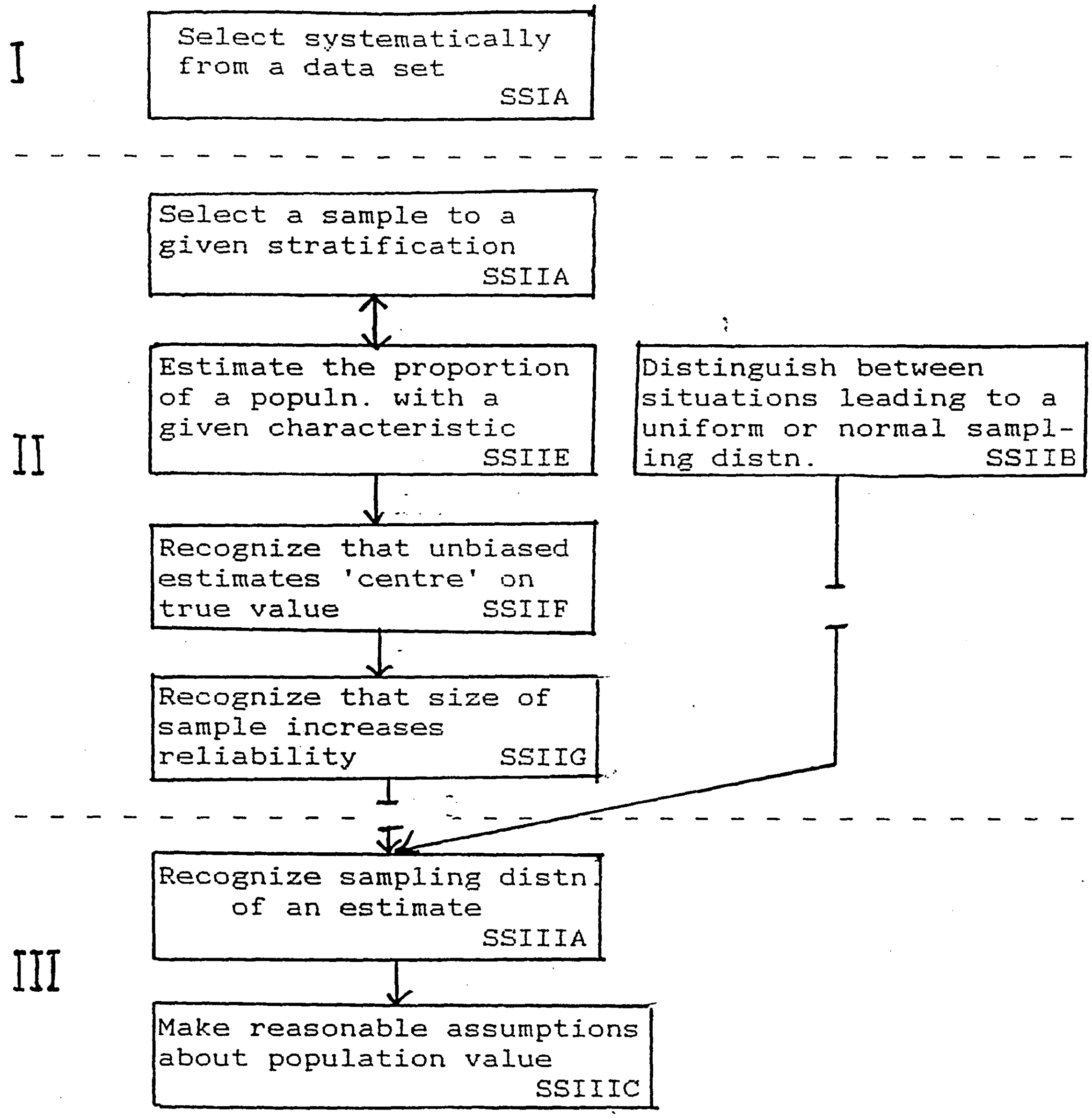
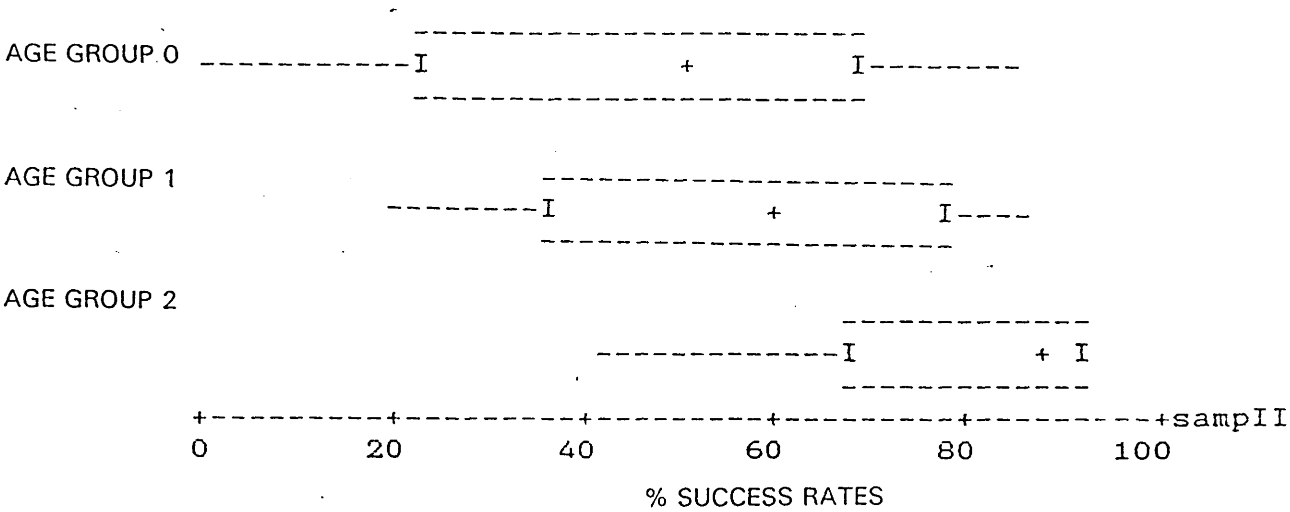
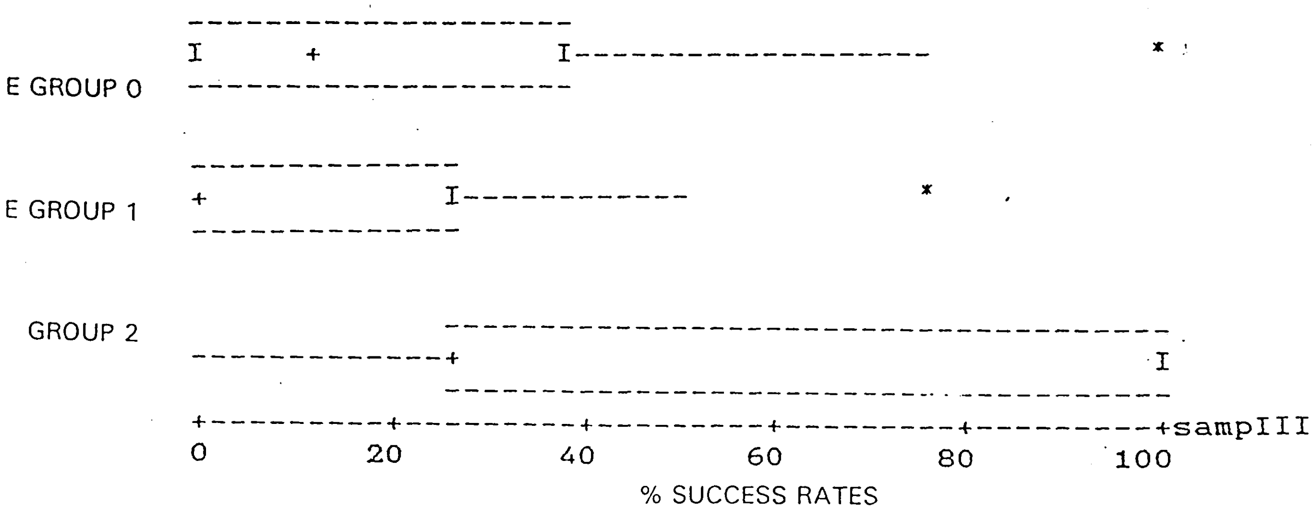


Figure 6.3.5.(i) Final hierarchy showing linkages in Sampling Skills



Correlation of % success at level II and Age, r = 0.475



Correlation of % success at level III and Age r = 0.202

Figure 6.3.5.(ii) Plots of % success at Level II and Level III against age in Sampling Skills

Success rates for the three age groups are shown in Table 6.3.5B:-

Table 6.3.5B <u>Percentage success rates at each level for different age groups</u>

	Maximum Level Achieved		
AGE	0/1	2	3
12.0-13.9	75%	15%	10%
14.0-15.9	76%	18%	6%
16.0+	20%	40%	40%

Patterns are less well defined here though in view of the fact that Level I contained only one item care must be taken at this level. It appears that the Level III skills were rarely achieved until the post-sixteen stage and Level II skills were little in evidence before this stage. The skills in this area are mainly those which will evolve by practical experience in dealing with samples, but in view of the project work carried out across the curriculum at GCSE it is surprising that a higher level of skill does not exist in the 14-16 age group.

6.4 Overall Patterns

The practicalities of administering the tests meant that, apart from a common group sitting Sampling and Distribution and Dispersion Skills tests, subjects took tests in only one skill area. This made direct comparison of the level of skill acquisition, and particularly the success at each of the three levels, difficult to compare across different areas.

Note has been made already of the general pattern which tentatively emerged in most skill areas at Level II. In looking at the level of success achieved by each age group at this Level the rates followed the same upward trend. That is, at the younger ages subjects vary greatly in the skills they have acquired, but older subjects are more likely to achieve a higher rate of success. This seems to justify a developmental learning process in that some young children are able to achieve success at higher thinking levels, but as age increases these higher levels become more accessible to all subjects. The children at the younger ages would not have met the higher skills within their normal teaching framework and thus their ability to complete these items suggests that they are derived concepts rather than taught ideas.

Comparing the percentage success rates over the different skill areas however shows some differences. In Sorting & Grouping the majority of subjects at the lower two levels achieved success at Level II and in Bivariate Data skills a greater percentage achieved Level II. Table 6.4.1 shows the overall numbers at each level:-

Table 6.4.1 Percentage success rates at each level by subjects of all ages

	Maximum Level Achieved		
	0/I	II	III
Sorting & Grouping	18%	61%	21%
Location	80%	14%	6%
Dispersion	73%	23%	4%
Bivariate	50%	31%	19%
Sampling	64%	21%	15%

Figures 6.4.2 (i),(ii) and (iii) show the percentage success rates of subjects in each of the three defined age group for each level in all the skill areas. The pattern above is shown in all skill areas, i.e. as a subject matures s/he is more likely to acquire the skills at Level II and Level III, though the youngest subjects are not precluded from acquiring these higher skills. This tends to confirm the notion of the cognitive acceleration theorists that age is not a

limiting factor in attaining these higher mental processes (Adey 1988).

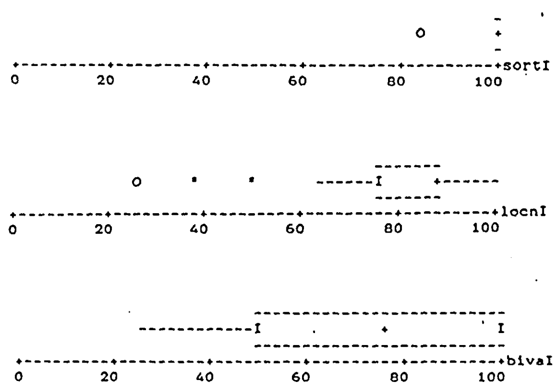
What is also evident from Figure 6.4.2 is the differences between skill areas in terms of the ages at which skills are acquired. The higher skills in Dispersion, and to a lesser extent Sampling, are rarely acquired before the age of 16. Higher skills in Bivariate Data and Sorting appear to be accessible to subjects at a younger age. Location skills show a progressively greater presence with increased age. These differences could be attributed to three factors:-

(i) Experimental error. With reasonably large samples and a fair degree of consistency in performance, as shown on some boxplots, this is thought unlikely.

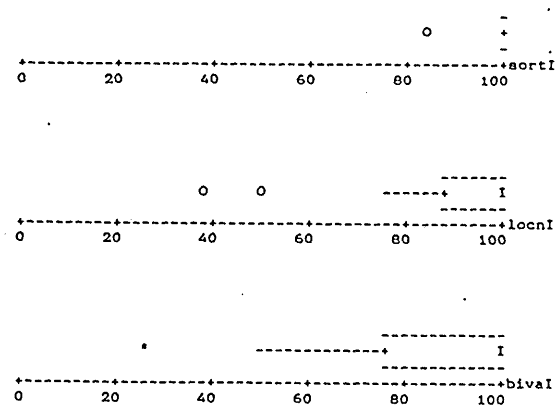
(ii) Poor definition of levels. Whilst considerable effort was made to classify skills in the correct level it is possible that particularly in Dispersion, some of the Level II skills are misplaced. Looking at the complexity of these skills, however, the skills appear to correspond to similar skills in Level II in other areas. The two skill areas of Dispersion and Sampling are areas not traditionally dealt with in pre-sixteen education and difficulties may have arisen as a result of unfamiliarity with the material.

(iii) Skill acquisition in different areas is not necessarily synchronized in terms of level. With more complex skills it may be that a wide variety of experiences are required before these can be acquired. These experiences may not necessarily happen to a subject on a time scale which enables skills to be acquired even though these are accessible. For example, to acquire the necessary sampling skills may require work in other subject areas on data collection or media exposure to surveys and sampling.

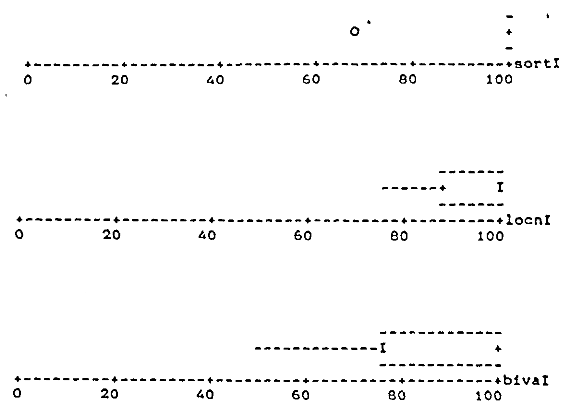
In order to establish the reason a detailed investigation involving a sustained programme of teaching and controlled experiences would be needed.



Age group 12.0 - 13.9

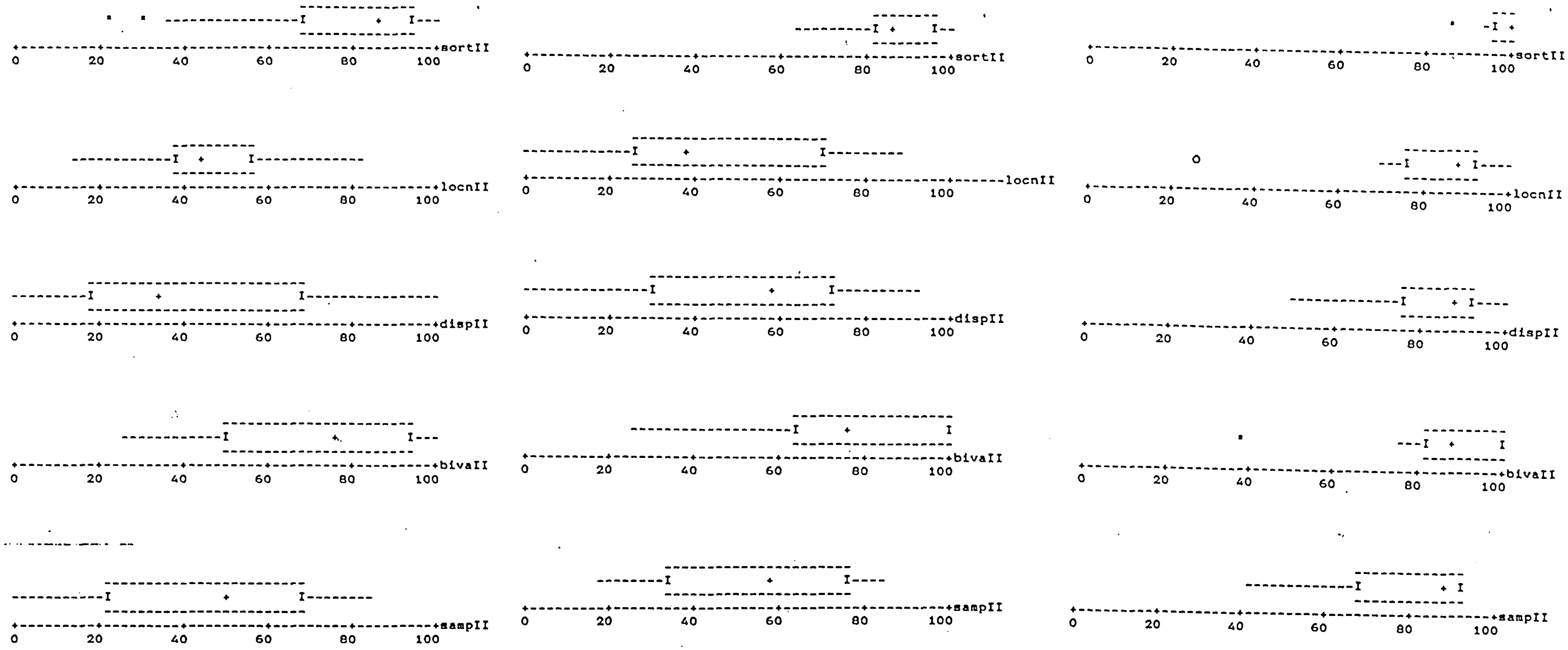


Age group 14.0 - 15.9



Age group 16.0+

Figure 6.4.2(i) Scores of subjects in different age groups at Level I in all skill areas

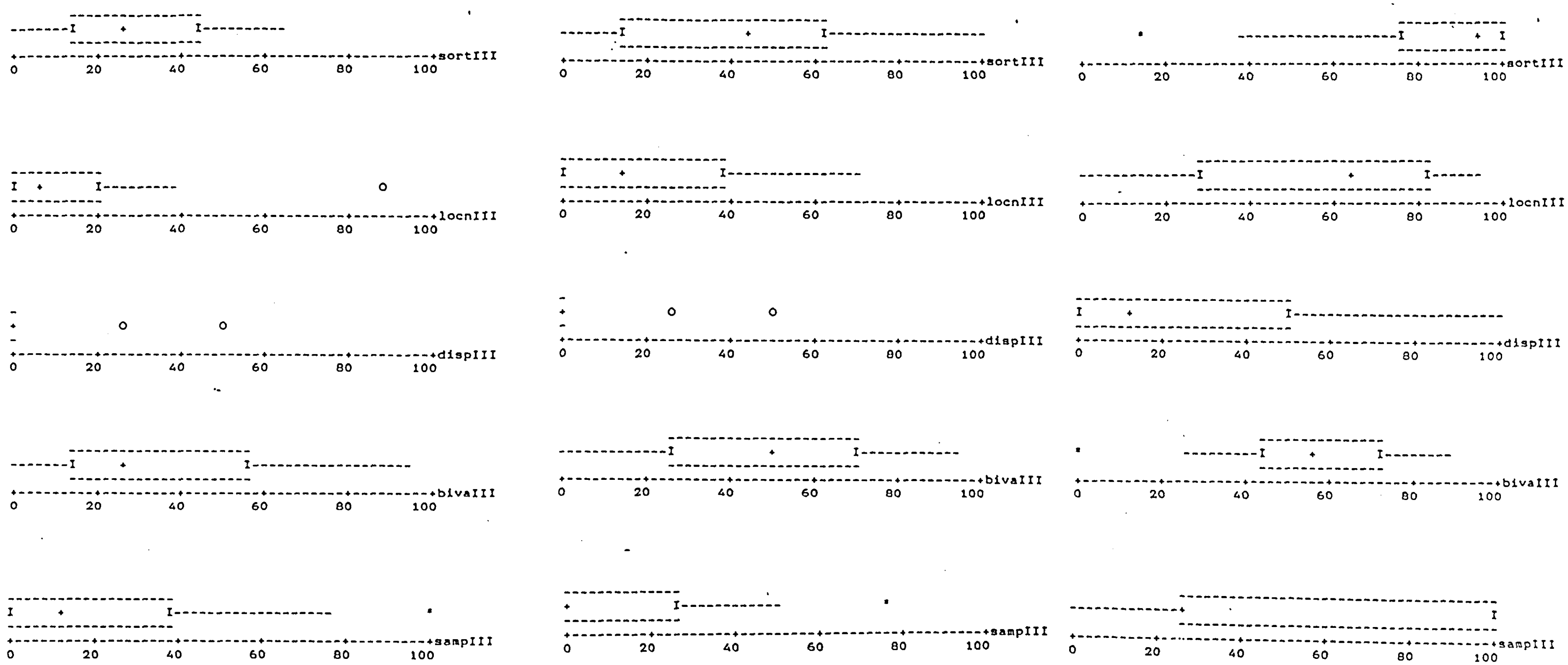


Age group 12.0 - 13.9

Age group 14.0 - 15.9

Age group 16.0+

Figure 6.4.2(ii) Scores of subjects in different age groups in all skill areas at Level II



Age group 12.0 - 13.9

Age group 14.0 - 15.9

Age group 16.0+

Figure 6.4.2(iii) Scores of different age group subjects in all skill areas at Level III.

6.5 Gender Differences

Subjects were asked to indicate whether male or female on the front top hand corner of the test paper. This was largely to ensure that an approximately equal number of males and females were tested at each level. Analysis showed this to be within +/- 4% of an equal distribution at each of the three age ranges. Some concern was felt as to whether in the highest age group there were disproportionally fewer girls taking mathematics at a higher level. Investigation showed a slight bias but as this was a small group anyway this did not appear to be significant.

Carrying out a full analysis, on the lines of the previous work, separately for male and female was not possible because of the sample size. Tests were carried out on the mean percentage success rates for boys and girls at each age group and for Levels where it was meaningful. Again where subjects all had a high level of success (e.g. Sorting & Grouping Level I) or a very low level, useful analysis was not possible. In those areas where it was possible to test, no significant differences were found between male and female performances.

6.6 Conclusions

The major hurdle in this research has been the range of subjects used for the testing. By the age of 12 it appears that many basic statistical concepts have been acquired already. This not only prevented the order of acquisition of skills being distinguished but also rendered any investigation of the age of acquisition difficult. Similarly at the upper end, in some areas where there was only a small number acquiring the skill it was difficult to make order distinctions. At one point it was considered giving only older subjects questions at the higher level over all five topic areas and conversely at the lower end. This would have enabled more questions to be set on each skill and a more detailed Inclusion Analysis to take place. However in view of the wide range of abilities, particularly at the lower age range, this would have hampered the ability to look at the age of acquisition of skills.

The initial testing had proved extremely valuable in removing questions from the tests which were inappropiate. The final tests on the whole gave very few problems in their design. The lengths gone to make the tests as visual and 'user friendly' as possible seemed from observation to have given subjects a high degree of motivation and encouragement.

In examining the order of acquisition of skills it was felt that some measure of success was achieved.

After some re-arrangement of initial skills the final hierarchies went largely uncontradicted. Whilst it was not possible to confirm the existence of some of the precedents many came out highly significant. Various key elements emerged in individual topic areas:-

(i) <u>Sorting & Grouping Skills</u> - The ability to sort data into categories is clearly a skill acquired early in a child's development. Even to the extent of using different forms of data in frequency tables this presents little problem to children by the early teens. Although it was not possible to tell whether the skill of compiling frequency tables is possible without necessarily being able to interpret the results, the latter still seems to be acquired at an early age. This is however an area that all subjects are likely to have practised at Primary level. In order to progress further in this area, effectively from putting data into given categories to making up their own categories, subjects need an understanding of the nature of more precise groupings. Only those in the oldest age ranges, with presumably greater experience, were able to understand the precise group definitions required and the implications of different groupings in order to set up and use their own tables competently.

(ii)<u>Measures of Location</u>:-This was the most complex of the hierarchies and although many precise conclusions of precedents were not possible, a skeleton pattern emerged. The earliest skills centred around finding measures where the answer was a whole number and, in particular, one of the data set. The mean is rather a more difficult concept which seems to be understood in the context of its balancing properties. One of the biggest stumbling blocks in all strands appears to be the using of an answer which is 'non-real' i.e. either where the answer is not one of the data set or, even more difficult, not a possible concrete example. The classic example being the difficulty of understanding what is meant by 'the average size of family is 2.2 children'. Finding the values of means and medians from frequency tables and weighted means involved much more complex skills and few could complete these. What is perhaps of more concern is the lack of understanding of the limitations of the various measures by even the oldest of subjects. This leads to the conclusion that even at the end of many years of experience of the measures most people have only an algorithmic knowledge of measures of location.

(iii) <u>Distribution & Dispersion</u>:- This was a more limited area than the others and the analysis was less conclusive. Many subjects at the lower and middle age ranges have a poor concept of the overall idea of spread. Even the idea of a simple range seems to be only partially understood by many. As age progresses the concept of spread becomes more likely to be related to the idea of centrality than homogeneity amongst the data. The more difficult ideas of quartiles and presumbably other measures of dispersion are apparently only attainable by the oldest subjects.

(iv)Bivariate Data:- This was the most involved hierarchy as there were in it strands between which there was little direct connection. Much of the initial hierarchy in this area was radically altered and of all the areas this perhaps gave the most enlightening insights into the way concepts in the topic are understood. The topic was also of great interest as many of the ideas here are not ones which the younger ages would have met in formal teaching.

The basic idea of 'correlation' is not a difficult one for even the younger subjects. Suprisingly, they even seemed to cope well with the idea of negative correlation. The use of bivariate frequency tables with qualitative and discrete data were well understood by a majority and enabled fairly quick progression onto dealing with the grouped situation. Despite this there seemed to be more difficulty in understanding what these tables actually showed in terms of variable relationships. The nature of this difficulty is clearly something worthy of further investigation. Plotting paired data on a scattergram rarely proved difficult, though many found obtaining a straight line from this rather harder. Without having developed these skills subjects seem unable to progress to more sophisticated judgements about correlation.

(v) <u>Sampling Skills</u>:- After the considerable simplification of the structure this was the most strongly supported hierarchy. This perhaps suggests that simpler hierarchies in other areas might have been useful. The 'pinnacle' of this hierarchy was in a crude way to understand the basic principle of a confidence interval, i.e. an estimate of a parameter is probably close to the true value. The series of skills needed to acquire this was strongly supported by analysis. The increasing understanding of the skills with age was also most marked with only a few of the oldest subjects obtaining full understanding. This an area rarely dealt with in school curricula and possibly shows the clearest example of skills acquired through everyday experience.

Whilst the precise linkages might have been difficult to analyse in some areas there is nevertheless strong support for the existence of three levels of concepts. Whilst these may not concur in the five topic areas consistently with age, the general pattern of development seems consistent. Indeed one of the difficulties in analysing the precise linkages was due to the fact that in most areas all the subjects had largely acquired the Level I skills and only the most advanced had acquired the Level III skills. The structure used by Green[1982], Fischbein[1975] and others, as outlined in Chapter 2, of the three stages of development seems consistent with the results obtained.

Chapter 7

Implications for Curriculum Designers and Teachers

7.1 Background

At the outset of this project it was hoped to gain a greater knowledge of the way in which statistical understanding developed amongst 12 - 18 year olds. What were the preconceptions they brought in from earlier experiences, what were the stumbling blocks to understanding and what were the limitations of understanding for children at different ages? Although many questions remain unanswered, the research has nevertheless given many insights into the development of statistical concepts.

Statistics as a taught subject in the UK has been an area of greatly increased interest in general mathematics teaching, in 'user' subjects and as a subject in its own right. The Cockcroft Report in the 1980s had considerable affect on mathematics syllabi for pre-16 courses and the Schools Council Project on Statistical Education in its turn influenced the compilers of the Cockcroft Report in their strong recommendations to promote the teaching of Statistics. Not only did this lead to the large Data Handling section in the National Curriculum (1990) but also led to recommendations on teaching styles:-

" Statistics is essentially a practical subject and its study should be based on the collection of data, wherever possible by the pupils themselves. It should consider the kinds of data which it is appropriate to collect, the reasons for collecting the data and the problems of doing so, the ways in which the data may legitimately be manipulated and the kinds of inferences which may be drawn."

(Cockcroft Para. 776)

The result of these reports has been, via the National Curriculum, to ensure that statistics has a place in all school syllabuses for pre-16 education.

In addition to this there are now many subjects at GCSE and A Level (and beyond) where a large element of statistics is required. Traditional syllabi such as biology and geography have for some years, both in class exercises and in project work, required an understanding of basic statistical techniques. In more recent years new school subjects such as psychology, business studies and physical education at A Level all contain a substantial element of statistics. At GCSE level with increasing reliance on coursework this has meant in many subject areas an increased used of statistics. Both at A and A/S level courses in statistics have shown a steady increase in the last ten years against a general pattern of declining numbers taking mathematics at this level.

The overall pattern then is of a dramatic increase in the importance of statistics in the secondary school curriculum, much of the material being taught by non specialists who, often by their own admission, have only a mechanical knowledge of techniques. Although there have been many accompanying statistical texts for courses using statistics these tend to use an algorithmic approach to techniques and pay scant regard to a development of understanding of the concepts and principles behind them. Traditional texts on A level mathematics and statistics, with a few exceptions e.g. Practical Statistics (Rouncefield & Holmes, Macmillan 1989), use an algebraic and/or algorithmic approach with understanding promoted by proofs and in many cases techniques are presented with no regard paid to the underlying principles. The Associated Examining Board's A Level Statistics has required formal proofs as part of the examination process until very recently. Encouragingly, new and revised examination syllabi are becoming less reliant on an algebraic approach and the use of project work is becoming the norm. GCSE texts, particularly those from the SMP group, are rather more enlightened and tend to pay far greater attention to the development of concepts. The development of means in the SMP 11-16 booklets is a good example of this.

Many users of Statistics in other areas have a strong feeling that although they use techniques, they do not really understand how the technique works, why they are using it and what are the implications and limitations of the results. Texts written for these users tend to use purely algorithmic approaches. There seems then to be a need for those who design curricula, produce books and other materials and teachers themselves to use methods which develop concepts rather than encorage rote learning.

7.2 Recent Commentaries on Teaching Styles

With the increased interest in Statistics there has been much literature and work on teaching programmmes in Statistics. Abele [1989], drawing from many years work in the German system, laid down what he maintains as a 'proven method of learning':-

(i) Learning through experience: pupils solve problems carry out selected assignments, analyse questions and perform appropiate experiments.

(ii) Concept formation: pupils look for shared and essential characteristics and begin to generalize.

(iii) Representation: pupils use graphic or symbolic representations and learn to use more formal methods.

Much of Abele's work concentrates on slow learners and he emphasizes the value of learning through discovery. His suggested programme of courses relies on pupils collecting their own data and rearranging and representing it appropiately; looking at different measures and characteristic values and assessing their appropiateness; and finally developing more complex processes and critically examining existing ones.

Nitko & Lane[1990] propose a similar structure to the learning process but stress the need for assessment to be built into the teaching programme. They suggest using computer assisted learning which would assess concept formation as the student progresses. They do suggest however that 'Knowledge of conventions, symbols, and systems underlie successful problem solving activity'. Experience from this project however suggests that many concepts and the ability to find solutions have little reliance on symbolic representation.

Several projects of a more practical nature have been developing new material for schools in recent years. The project 'Statistical Investigations in the Secondary School' based at the Open University and DES funded set out to introduce some of the new methods in statistics into the secondary school curriculum and develop work using microcomputers. The materials focussed on three main areas:-

(i) use of graphical methods

(ii) principles of a statistical investigation

(iii) use of microcomputers to develop the above.

As a result of their work they put forward a teaching programme where pupils collect their own data and then used a micro-computer to carry out the sorting processes and graphical representations. The pupils could then devote their time to interpreting their results and assessing the procedures according to the first two processes laid down by Abele.

In the US the 'Quantitative Literacy Project', carried out by the National Research Council [1989] had similar aims. The only difference was that the target was to provide supplementary material to existing courses. Again in this project emphasis was laid on pupils collecting their own data and analysing the results.

The 'Project on Statistical Education' in the early 1980's started by looking at the needs of non-specialist young workers and statistical requirements in the world at large. Materials were then produced using examples from many of the user subjects such as Biology and Economics and examples from the commercial world. Another project from Sheffield 'Practical Work in A Level Statistics'(1989) trialled and developed a set of materials for work at the higher level. The underlying principle of the book, now published, is of collecting data from a practical experiment looking at ways of examining the data and using this to evolve a formal technique. Although this was aimed at A level many of the ideas are usable at a lower level and the approach fits in well with the pattern of learning structure described by Abele and supported by this project. The most recent project from Sheffield 'Handling Data Using Spreadsheets' (1991) is considering using data on spreadsheet packages to develop techniques.

These projects all recognise the growing need for new courses in Statistics which use an approach which develops concepts rather than teach as an algorithmic technique with minimal understanding. In particular the necessity for pupils to gain experience from handling data in real life situations is becoming fundamental.

7.3 The National Curriculum & Topic Inclusion

The most significant guiding force in pre-16 Mathematics education in the UK at present is the National Curriculum. Two out of the original 14 targets are devoted to 'Handling Data' (Appendix 1) stressing the importance which those designing the curriculum felt statistics deserved. The original 14 Attainment targets have now been condensed into five National Attainment Targets, one of which is entirely statistical and probablistic material. Although on its introduction in the late 1980s much of what it contained was already present in most modern mathematics courses there were some interesting developments, particularly with regard to the Levels at which topics were to be introduced. The use of scattergrams and subsequently drawing best fit lines, interpreting two way tables and generally handling bivariate data were notably included starting at Level 6.

For the purpose of this work the National Curriculum fits very neatly into the three age categories used in analysis. Despite the overlapping of the levels the Key Stage divisions 2,3 & 4 give a good indication of what topics pupils should be working on within the age bands. The table below gives the Key stages, levels covered and the 'reporting' age (i.e. the age by which all pupils should have been introduced to topics), and a summary of the topics covered in this research included in those levels within that stage.

KEY STAGE & LEVELS	REPORT AGE	TOPICS
KS2 Level 2-6	11	Sort & classify qualitative data and put into frequency tables. Collect, group and order discrete & continuous data and construct and use frequency tables. Interpret frequency tables. Understand, calculate and use the mean & range of a data set. Qualitative bivariate tables. Create scattergram and have basic understanding of correlation.
KS3	14	Calculate the mean, median and mode from a frequency table. Construct a cumulative freq. table and curve. Use it to find median & quartiles. Draw line of best fit by eye on a scattergram.

KS4	16	Describe the range of a variable using other measures of dispersion(including S.D.). Consider different shape distributions. Use sampling and recognise effect of size on reliability.

Whilst this should not be regarded as a rigid list of what a pupil should know and understand at a particular age, it nevertheless provides a guide to the topics which are expected to be introduced at each stage. The Department of Education & Science literature is clear however that not all pupils will be expected to have studied all the topics, in particular Key Stage 4 topics would only be tackled by more able pupils.

From the work in this project it is possible to examine the topic areas in the light of the difficulties experienced and the ages by which pupils seemed to have acquired various concepts. Key Stage 2 of the National Curriculum will be reported on at the age of 11, therefore the topics in this area should be, at least, familiar to all pupils of secondary school age. The work on Sorting & Grouping Skills shows that few pupils have any difficulty with basic sorting and use of frequency tables. However what was revealed was the difficulty pupils had in compiling their own grouped frequency tables, particularly with continuous data. Finding the mean and range of a data set was difficult in itself. What was apparently lacking in most pupils however was a real understanding of what the mean in fact showed. In particular the problems encountered explaining the significance of a non-real answer till rather later is clearly a stumbling block. There must inevitably therefore be large numbers of pupils who are able mechanically to calculate the mean with limited understanding. In view of the high correlation of understanding of the mean and ability to carry out the balance exercises it is recommended that work of this nature be used in connection with the mean to promote understanding.

The use of simple qualitative bivariate tables seems applicable at this level though there appears to be some difficulty in interpreting their meaning in certain cases. Scattergrams present few problems. Whilst the basic idea of correlation is understood, however, from casual observation of data or from bivariate tables, this is not evident from a scattergram unless the pupil is able to visualize a straight line in the data . The research seems to indicate that this is not as easy as is often thought. Whilst there is no mention of negative correlation in the National Curriculum it would seem reasonable to include in any discussion of correlation. Texts however tend to ignore it. The evidence in this project suggests that it is not a difficult idea for even the youngest tested and needs to be developed for a fuller understanding of the concept of correlation.

Key Stage 3, with its reporting at age 14, requires that all the topics therein are covered before the 'GCSE' years at 14-16. Even with broader guidelines certainly by the sitting of GCSE these ideas should be familiar to all. Calculating means and medians from frequency tables is a difficult idea and only possible for the more able. This project tends to suggest that more in depth work with frequency tables is required at an earlier age and that working from frequency tables is a fairly abstract concept. To an even greater extent the use of cumulative frequency tables causes great difficulty for all but the most able, in particular using them to find quartiles. Mention has already been made of the difficulty that pupils have of visualising best fit lines on a scattergram. This is not only an essential skill in understanding correlation but has other implications in the ability to quantify relationships between variables particular in user subjects such as Science.

The Key Stage 4 topics are supposed to be covered during the GCSE course under the non-statutory guidelines, though it is pointed out that many will not be expected to have met the later stages (i.e. 9 & 10). These topics are basically represented by more advanced dispersion measures and recognition of certain distribution patterns. Certainly few GCSE courses deal with standard deviation and whilst this was excluded from the main testing of this project the initial testing indicates that this is really something to be covered in A level courses. The whole concept of dispersion is one that appears not to be developed until the middle years of secondary education. The recognition of different distributions in terms of sampling was only possible for those in the 16+ age group with any degree of success. Understanding of the concept is probably only evolved over a period of time with exposure to many data sets of different types.

One area that seems to have been largely ignored by the National Curriculum until very late is the development of skills which lead to an understanding of the problems of sampling from a population. Much work is carried out in lower levels on questionnaire design and pupils are at an early stage expected to carry out surveys and otherwise collect numerical data. Without an awareness of the problems of sampling, practical work, particularly in coursework for GCSE, will be very much weakened. Much of this area seems intuitive to those in the middle years and could easily be introduced at a lower level.

7.4 Statistics Curriculum in the United States

In the United States, though with much less Government intervention, a similar development has occurred in Mathematics teaching. The report 'Everybody Counts', National Research Council[1989], mirrors much of the work of the Cockroft Report in its examination of the needs of school pupils for Mathematics in the outside world.

> "Most mathematics should be presented in the context of its uses, with appreciation of mathematics as a deductive logical system built up slowly through the levels of education."

It lists too, as one of four areas deserving of greater emphasis "exploratory data analysis and statistics which facilitate reasoning about data".

This report in turn contributed to the production by the National Council of Teachers of Mathematics the publication in the same year 'Curriculum and Evaluation Standards for School Mathematics'[1989]. Whilst non statutory the influence of NCTM is great and the standards are likely to have a significant effect on the teaching of mathematics in the US. The 'Standards' lists topics in the same way as the National Curriculum with slightly more detail on the actual implementation.

Direct comparison with the National Curriculum is slightly confounded by the incompatabilities of the age structure, but nevertheless the work in this project can be directly related to the Standards. In the first stage taking pupils up to Grade 4 (i.e. age 9/10) Standard 11: Statistics & Probability centres around collecting, organising and describing data including the use and criticism of diagrams. The work here on Sorting & Grouping, whilst not having been tested at the appropiate ages, suggests that this is suitable and indeed sound work at this level. In the next stage up to grade 8 (Year 9 in English schools) this work is extended to include 'averages' and range at a comparable level to the National Curriculum though a greater emphasis appears to be laid on the comparison of the mean, median and mode and their interpretation. A notable inclusion at this level is the discussion of types of sampling and the effect of size of sample on precision.

At the next stage, for those at High School up to age 18, the Curriculum and Evaluation Standards suggests a number of new statistical ideas should be introduced, even for those not 'intending to proceed to college'. Concepts on bivariate issues such as correlation, best-fit lines etc. are dealt with for the first time, rather later than their UK counterparts. Measures of dispersion are first introduced on a parallel with the National Curriculum. What is again significant is the development of ideas in sampling.

On the whole the US Standards give a similar, if not greater, emphasis to Statistics. Commentary in the Standards makes it clear that a study of Statistics at High School is essential for coping with modern life. The emphasis is very much on handling real data and carrying out experiments with due regard to the methods of sampling. As with the UK, emphasis is put on using databases but their ideas go even further in encouraging the use of software to enable greater exploration of techniques.

7.6 Suggestions for improvements in new courses

As yet the separation of taught and intuitive ideas has not been discussed. In the testing the intention was, as best as possible, to explore what was intuitively learnt rather than what or how it was taught, though inevitably much content had probably been taught, particularly in the first two topic areas. With a few reservations the testing indicates what is a reasonable pattern of development of concepts and at which ages it seems appropriate to attempt to introduce these concepts. It could be argued that, even where teaching had taken place, the failure to succeed in items in this test indicates a lack of understanding of the underlying concepts.

Clearly Sorting & Grouping skills are fundamental to progression in statistics. What has been found however is that frequency tables are not always fully understood and progression to using these to calculate measures, such as the mean, is severely inhibited by this. In addition many pupils find it difficult to set up their own grouped frequency tables. Compiling and interpreting group frequency tables is clearly essential in coping with real life situations, and it is believed that only through handling real data can these skills be developed. However with the introduction of stem and leaf diagrams doubt exists as to whether frequency tables and their compilation is an essential tool. Stem and leaf diagrams have many of the properties of a frequency table yet the pupils can more readily visualize the original data. The use of cumulative frequency tables was found very difficult too. It could again be argued that stem and leaf diagrams render this slightly superfluous and the ability to distinguish exact data items should make them more concrete and avoid the assumptions of continuity required in the cumulative frequency diagrams.

The main difficulty emerging in the work on location measures was the problems with 'non-real' answers, i.e. the median of an even quantity of data and a mean of discrete data not itself discrete. In the former case the use of stem & leaf diagrams might again promote a better understanding of the 'middle' of a data set. The link between the mean and balances has been mentioned already and the need perhaps to evolve the mean from practical work with balances. A useful intermediate stage is the use of 'dotplots' now used in SMP texts and drawn by many computer packages. These, like the stem and leaf diagrams, allow complete data to be visualized and look remarkably similar to the balance diagrams used in the tests. The effect of outliers in particular may be evident once the effect on balances has been understood.

Work on bivariate data, whilst now being included to a greater extent than previously, could be developed still further. The research has shown that the basic ideas of correlation, scattergrams and bivariate tables are understood at an early age. The extension into fitting straight lines clearly needs more emphasis and work on this in the middle years of secondary school is recommended. The extension of bivariate tables for the more able would be a good development of the work on correlation and the use of ranking could usefully be developed at an early stage.

In the cases of location measures and bivariate data the use of microcomputers is encouraged. The ability to visualize data in the form of stem and leaf diagrams, dotplots, bivariate tables, scattergrams and best fit lines is clearly important. Using packages such as MINITAB to execute the more tedious sorting and graphical applications would enable pupils to handle larger sets of data and gain a far greater variety of experiences on which to judge the merits of different techniques such as the mean. Indeed the use of real data collected by pupils should be used whenever possible. In the work on correlation, subjects were able to gain intuitive ideas about correlation simply by looking at a raw set of data and the fact that these intuitive ideas can be supported and checked by various techniques should enable these concepts to be readily reinforced.

The area where it is felt that most change needs to be made is in the development of ideas of sampling along similar lines to that in the US standards. Methods of selection of samples can be dealt with at an early age and need to be in order that any discussion of the results from practical work can be carried out. This would then facilitate work on the nature of the 'representative' and 'variability' principles, as laid down by Rubin et al.[1990], to be developed. This would link in with current work in the National Curriculum on different distributions to provide the necessary groundwork for the later ideas of statistical inference. The many testing procedures used in the 'user' subjects at A level which contain a large statistical content desperately need these ideas if techniques are not to become a meaningless set of algorithms.

Chapter 8

Conclusions and Suggestions for Further Work

8.1 Critical Analysis of the Project

The inspiration for this project arose from the realisation that apart from some work on 'averages' very little work had been done on the way children's concepts in statistics developed. It was hoped therefore to give a better picture of the pattern of concept development, the problems which inhibit the development of higher concepts and to see if this development was in any way restricted by age considerations. The results could then be used to examine current curriculum and teaching methods and suggest ways in which these could be improved.

Initially it was hoped to carry out tests over a wide range of schools but practicalities prevented this. It was felt however that because of the truly comprehensive nature of the researcher's own school that a reasonably representative sample was in fact obtained. With more than 20 initial feeder schools and a 60% external entry into the sixth form subjects had been exposed to a range of previous teaching experience. The sample size, whilst allowing for representativeness, caused limitations in the analysis able to be carried out. The major restriction on the extent of the findings in this project was the limitations superficially imposed on age. Many of the earliest intuitions of statistics had been developed prior to the secondary school age and testing of the development of these early concepts was not possible. Had the project been extended to younger children a review of the testing procedure would have been necessary as language would have caused problems in the written test format. Similarly, though less significantly, testing development of the higher concepts was limited by the small number in the sample who were able to achieve this level.

Some doubt has been already expressed in the literature as to the validity of rigid learning hierarchies. Whilst many specific links were not proven a pattern of key concepts emerged and there was general support for a progressive learning model, in particular the three levels of concepts as defined earlier. The establishment of hierarchies and their testing has not been widely used as a research technique, and there is no evidence of its use in this subject area. The Clarke & White Inclusion Analysis technique is rarely included in literature and whilst it is not a particularly robust method it can nevertheless be a useful indicator of learning patterns. Suggestions have been made in the literature that learning patterns vary between individuals but it is difficult to see how without very large samples this can be carefully tested with due regard paid to experimental errors.

Much time and effort was devoted to the design of the written tests. It was the intention that by using visual material, simple language, and careful question design the written tests would be easily accessible to all the participants. Whilst there were still some difficulties over question design, the overall format of the tests was very well received and failure to complete the test was virtually nil. In particular, the highly visual data sets made the tests attractive and stimulated a great deal of interest amongst participants and supervisory staff.

The insight which the testing gave into a pupil's understanding of statistics, whilst by no means providing a complete model, gave useful information to examine current curricula and teaching practice. Some of the difficulties currently experienced were identified and areas which are currently neglected were able to be promoted for future course design. In particular, the desirability of more work on bivariate data early in the secondary years needs to be recognized as well as the inclusion of work on the problems of sampling.

8.2 Suggestions for Further Research

Inevitably a study of this nature throws up more unanswered questions than those it solves. This project by its nature covered many areas in a general way and there is a great deal of scope for further work in specific areas.

As stated earlier, definition of precise links at the early levels of concepts was inhibited by the age restriction of subjects who were of secondary school age. Further work would require rather different forms of testing with children younger than this. The visual nature of the data sets could be employed though the material could be presented in an interview situation rather than as written tests since language problems could interfere with the true test of concept acquisition. A further step would be the use of physical objects: sorting actual objects; making judgements of centrality, dispersion and association; selecting items from a large data set; using balances. Again taped or videoed interview techniques should be employed.

At the upper end of the hierarchies the tests would be equally applicable. By using selective samples of older students questions at the lower levels could perhaps be ignored, simply testing some key 'minimal entry' skills, and the skills in question could be examined more carefully.

One theory which tentatively emerged from the work was the idea of 'key concepts'. There appear to be certain skills which inhibit further learning until they have been acquired. Whilst some of these may have already been identified there are perhaps others which the structure used did not identify. Identifying these, the extent to which they inhibit progression, any apparent age barriers and how these key skills can be more readily acquired would be valuable areas for further research.

Much work has been carried out already in the area of measures of location, being a key area in statistics. One particular area which was not tested and requires perhaps different procedures is to look at children's instinctive use of the three measures in different situations. Presented with a variety of different situations when would they choose a median, mode or mean as the 'representative value' of a data set and does this choice change with age? Some initial work has already been carried out with coloured balls and rods along these lines by the researcher. Work on measures of dispersion could be evaluated as well, but as this is a later concept in the general pattern of concepts this is rather more difficult. A similar 'strategy' style approach could be used in investigating children's natural intuitions regarding methods of sampling. A brief analysis was possible in this project based on errors, but this is an area worthy of further investigation.

Of all the areas which were perhaps inadequately tested by this project Bivariate Data is probably the most prominent. Since this is an area fairly recently introduced into the normal school curriculum, and it is recommended here as being worthy of greater inclusion, it certainly warrants further investigation. In the hierarchy given for this area there appear to be four major strands; bivariate tables, scattergrams and lines of best fit, correlation and ranking. There is much interlinking of strands and the picture is perhaps more complex than has been suggested here.

With testing of Key Stages of the National Curriculum being introduced in the next few years detailed information on general ability in statistics should become available. Analysis too of the appropriateness of topic inclusion at particular levels could be made. As teaching material continues to be developed in line with new curriculum change and new approaches it will perhaps be possible to study further some of the issues raised in this project.

REFERENCES

ABELE A.(1987) How to Teach Statistical Concepts to Slow Learners, Proceedings of ICOTS II(Univ. of Victoria)

ABELE A. (1989) How to teach statistical concepts to slow learners, Studies in Mathematics Education(UNESCO) Vol. 7, 146-152

ADEY P. (1988) Cognitive Acceleration: Review and Prospects , International Journal of Science Education Vol.10, No.2, 121-134

ADEY P. & SHAYER M. (1992 Unpublished) An Exploration of Long-Term Far-Transfer Effects Following an Extended Intervention Programme in the High School Science Curriculum Manuscript for 'Cognition & Instruction'

AIRASIAN P., MADAUS G., WOODS E. (1975) Scaling Attitude Items: A Comparison of Scalogram Analysis & Ordering Theory, Educational & Psychological Measurement 35, 809-819

ASSESSMENT OF PERFORMANCE UNIT (1980) Mathematical Development - Secondary Survey Report No. 1 H.M.S.O.

BARR G.V. (1986) Some Student ideas on the Median & Mode, Teaching Statistics 2, 38-41

BEINS B.C. (1987) Student Knowledge of Statistics, Paper presented to American Psychological Association

BERGAN J. (1980) The Structural Analysis of Behaviour:An Alternative to the Learning Hierarchy Model, Review of Educational Research Vol. 50.4

BEYTH-MAROM R. (1983) Choosing an Appropriate Summary Statistic, Teaching Statistics 5.3, 70-74

BRUNER J.S. (1960) On Learning Mathematics , Mathematics Teacher (Dec 1960)

CAMBRIDGE TRUST FOR EDUCATION (1981) APU Tests - New Perspectives in the Maths Curriculum

COCKROFT et al (1982) Mathematics Counts - The Cockcroft Report, HMSO

CRESSWELL M. & GRUBB J. (1987) The Second International Mathematics Study in England & Wales, NFER

DIENES Z.P. (1967) A Theory of Mathematics Learning from 'Building Up Mathematics Learning' (Hutchinson, London)

DUNKELS A. (1987) EDA in the Primary Classroom - Graphing and Concept Formation Combined, Proceedings of ICOTS II (Univ. of Victoria)

FISCHBEIN E. (1975) The Intuitive Sources of Probablistic Thinking in Children, D.Reidel

GAGNE R.M. (1970) The Conditions of Learning(2nd Ed.),(Holt, Rinehart & Winston;New York)

GAL I., ROTHSCHILD K., WAGNER D. (1989) Which Group Is Better?: The Development of Statistical Reasoning in School Children, Paper presented to Soc. Research in Child Development

GARFIELD J. & AHLGREN A. (1987) Difficulties in Learning Probability & Statistics, Proceedings of ICOTS II (Univ. of Victoria)

GOODCHILD S. (1988) School Pupils' Understanding of Average, Teaching Statistics. 10.3, 77-81

GREEN D.R. (1982) Probability Concepts in 11-16yr Olds, CAMET

GUTTMAN L. (1944) A Basis for Scaling Qualitative Data, American Sociological Review 9, 139-150

HSU L.M. (1981) On Scales of Measurement, Teaching Statistics 2.3, 86-89

JOLLIFFE F.R. (1986) Two Sampling Misconceptions, Teaching Statistics 10.1, 66-69

KAHNEMAN D., SLOVIC P., TVERSKY A. (Eds) (1982) Judgement under uncertainty: Heuristics and Biases , C.U.P.

KATZ L. (1985) The Difference in Proportions as a Measure of Association, Teaching Statistics 7.2, 40-42

LEON M.R. & ZAWOJEWSKI (1990) Use of the Arithmetic Mean: An Investigation of Four Properties. Issues and Preliminary Results , Paper to 3rd Int. Conference on Teaching Statistics

LEWES F. (1981) Where should we start, Teaching Statistics 3.3, 66-71

LOOSEN F., LIOEN M., LACANTE M. (1985) The Standard Deviation: Some drawbacks of an Intuitive Approach, Teaching Statistics 7.1, 2-5

LOVIE P. & A.D. (1976) Teaching Intuitive Statistics 1: Estimating Means & Variances, Int. Journal of Maths Educn.in Science & Technology 7.1, 29-40

MEVARECH Z.R. (1983) A Deep Structure Model of Student's Statistical Misconceptions, Educational Studies in Mathematics 14, 415-429

MICK H.W. & BRAZIER G.D. A Logical Cognitive View of Learning and Teaching Mathematical Concepts Mathematics Education Report (ERIC)

MITTLEHOZ D.J. ET. AL. (1985) Learning & Intelligence: Ability differences in the organization and recall of elementary statistical concepts. Paper presented to A.E.R.A.

MOKROS J. & RUSSELL S. (1989) Whats Typical?, Hands On! Vol. 12, 8-9,21

MOSTELLER F., SIEGEL A., TRAPIDO E., & YOUTZ C. (1981) Eye Fitting Straight Lines, American Statistician 35.3, 150-152

NATIONAL RESEARCH COUNCIL (U.S.) (1989) Everybody Counts

N.C.T.M. (1989) Curriculum & Evaluation Standards for School Mathematics

N.C.T.M. Results from the 2nd Maths Assessment of National Assessment of Educational Progress

NITKO A.J. & LANE S. (1990) Solving Problems is not enough: Assessing and diagnosing the ways in which students organize statistical concepts , Paper presented at the 3rd Int. Conference on Teaching Statistics, New Zealand.

POLLATSEK A., LIMA S., WELL A.D. (1981) Concept or Computation: Student's Understanding of the Mean Educational Studies in Mathematics 12, 191-204

POSTEN H.O. (1982) Neat Data for Teaching Statistics, Teaching Statistics 4.2, 42-45

RESNICK L. (1973) Issues in the Study of Learning Hierarchies, Instructional Science Vol.2

RESNICK L. & FORD W. (1981) The Psychology of Mathematics for Instruction, Laurence Eblaum Assoc.(N.Y.)

RUBIN A., BRUCE B. & TENNEY Y. (1990) Learning about sampling: Trouble at the core of statistics , Paper presented at the American Educational Research Association.

SCHUPP H. (1987) Appropiate Teaching & Learning of Stochastics in the Middle Grades, Proceedings of ICOTS II (Univ. of Victoria)

SKEMP R. (1986) The Psychology of Learning Mathematics(2nd Ed.), Pelican

THORNTON Francis (1986) Pilot Survey for an Investigation into Children's Understanding of Graphics in Statistics, Sheffield Poly. (Unpublished)

WINTER H. (1983) Why teach descriptive statistics in Secondary schools?, Proceedings of 1st Inter. Conference on Teaching Statistics Vol. 1, 127-143

WHITE R. (1976) Research into Learning Hierarchies, Review of Educational Research 43.3, 361-375

WHITE R. (1974) The Validation of a Learning Hierarchy, American Educational Research Journal Vol 11, No.2, 121-136

WHITE R. & CLARK R. (1978) A Test of Inclusion Which Allows for Errors of Measurement, Psychometrika 38, 77-86

APPENDIX 1: The original National Curriculum Attainment Targets as issued by the Department of Education & Science, showing Targets 13 and 14.

NAT 2

a	Knowledge of whole nos. up to 1000
b	Decimal notation in money
c	Meaning of -ve whole nos. in context
a	Read write and order whole nos.
b	x by 10 and 100
c	Measure to 2 d.p.
d	Recognise and understand simple everyday fractions
e	Recognise and understand simple everyday percentages
f	Understand & use the relationship between place values
a	Use index notation for while nos
b	Use unitary ratios
a	Read write and order decimals
b	Relate fractions, decimals, percentages and ratio
a	Express a +ve integer as a product of primes
a	Use s.f. notation
b	Use index notation to express powers and roots

a	+ and - nos. up to 20
b	Solve x and / problems with a calculator (whole nos.)
c	Use table facts up to 5 x 5
a	Use table facts up to 10 x 10
b	Mentally calculate + and - of 3 digit nos. x and / 2 digit by 1 digit
c	Solve x and / problems with a calculator (up to 2 d. p.)
a	x and / 3 digit nos by 2 digit nos. manually
b	calculate fractions and decimals of a quantity
c	Mentally x and / 2 digit multiples of 10 by 1 digit x's of 10 (whole nos)
	Use negative nos. in context
a	Work out fraction and % changes
b	Calculate using ratios
c	Convert fractions to decimals and %
	Mentally x and / 1 digit multiples of 10 by 1 digit x's of 10 (decimals)
b	Solve x and / problems (any nos.)
	Use memory and brackets on a calculator
	Calculate with nos in s.f.
b	Substitute -ve nos. into formula
	Calculate with fractions

	Approximate to nearest 10 or 100
b	Understand remainders & know to round up or down
	Use estimation to check + and - sums
	Read calculator to nearest whole no.
	Interpret calculator rounding errors
	Use and refine trial & improvement methods
	Approximate to specified sf's or dp's
	Use estimation to check x and / sum
	Use earlier skills in context
	Use estimation to check answers of the right order

Use a range of metric units
Use appropriate measuring instruments & read scales
Make estimates based on familiar units
Understand the relationships between units
Make sensible estimates of a range of measures
Use scale in maps & drawings
Use common imperial units & know equivalents
Convert from one metric unit to another
Understand & use compound measures
Measure to appropriate accuracy
Recognise that a given measurement is in possible error of 1/2 unit

NAT 3

Explain number patterns & predict next no.
Use number patterns & equivalent forms to perform mental calculations
Recognise multiples of 2, 5 and 10
Apply strategies such as x 2 and x1/2 to explore number properties; equivalent fractions
Generalise in words patterns from various situations
Understand & use terms such as prime,square (root), cube (root), multiples and factors
Recognise no. patterns through spatial patterns
Generate sequences from simple instructions
Determine rules for generating sequences
Use spreadsheets etc to explore no. patterns
Use symbols to express rules of sequences
Use reciprocals and reciprocal relations
Explore complex computer generated no. paterns
Understand the relationship between powers and roots
Understand and use dis-proof by contradiction

Inputs & Outputs from simple function machines
Simple formulae or equations in words
Recognise x and / as inverses and use to check
Understand and use simple formula/equations in symbols
Express a simple function symbolically
Solve linear equations
Solve simple polynomials by trial & improvement
Use rules of indicies in algebra
Solve simple inequalities on a number line
Solve a range of polynomials by trial & improvement
Solve algebraically simultaneous linear equations
Manipulate simple algebraic expressions
Solve linear and other inequalities
Understand and use a range of formula & functions

Understand & use coords in the 1st quadrant
Understand & use coords in all 4 quadrants
Use coordinate plots to represent simple mappings
Draw and interpret graphs of linear functions
Solve graphically simultaneous linear equations
Use computers to generate various types of graphs & interpret them
Know the graphs of quadratic and reciprocal fns
Use graphs to solve linear inequalities

NAT 4

OAT 10

3	a	Sort 2 and 3-D shapes in different ways
4	a	Understand & use language of angle
	b	Construct simple 2 and 3-D shapes
5	a	Understand congruence of simple shapes
	b	Use angle properties of lines and triangles
6	a	Use angle properties of polygons inc quadrilaterals
	b	Recognise and use 2-D representation of 3-D
	c	Use computers to generate and transform 2-D shapes
	d	Classify and define types of quadrilaterals
7	a	Understand & use pythagoras
8	a	Use sine, cosine and tangent in 90^0 triangles

OAT 11

3	a	Recognise reflective symmetry in 2 and 3-D shapes
	b	Understand & use eight points of the compass
4	a	Specify location by +ve coords and by angle and distance
	b	Recognise rotational symmetry
5	a	Identify the symmetries of different shapes
	b	Use networks to solve problems
	c	Specify location by means of coords in all 4 quadrants
6	a	Understand & use bearings to define direction
	b	Reflect simple shapes in a mirror line
	c	Enlarge by a whole number scale factor
	d	Devise instructions for a computer to proc=duce shapes
7	a	Determine the locus of a point
	b	Enlarge by a fractional number scale factor
	c	Use coords in 3-D (x,y,z)
8	a	Understand & use mathematical similarity
	b	Understand & use vector notation

OAT 8

4	b	Find areas and volumes by counting
5	d	Measure and draw angles in degrees
8	a	Calculate using length area and volume
	b	Use formula for perimeter, area and volume

NAT 5

OAT 12

3	a	Extract information from tables and lists
	b	Enter and access information in a database
4	a	Specify data collection; collect and tabulate discrete data in frequency tables
	b	Understand and calculate mean and range
	c	Interrogate data in a database
5	a	Design and use data sheets to collect data
	b	Collect group and order continuous data in frequency tables
	c	Insert and interrogate data in a database and draw conclusions
6	a	Specify data collection;use an appropriate observations sheet, collate and analyse
	b	Design a questionaire (2 responses), collate and analyse results
7	a	Specify and test a simple hypothesis
	b	Record and organise grouped data into class intervals; claculate mean
	c	Find mean, mode, range and median of frequency distributions and interpret results
8	a	Design a questionaire (3 responses), collate and analyse results
	b	Construct a cumulative frequency table

OAT 13

3	a	Construct and interpret bar charts
	b	Create and interpret pictographs
4	a	Create a flow chart to sort objects
	b	Construct and interpret stick graphs for discrete variables
	c	Construct and interpret line graphs; interpret intermediate values
	d	Construct and interpret an interval frequency diagram (discrete)
5	a	Construct and interpret pie charts
	b	Construct and interpret conversion graphs
	c	Construct and interpret an interval frequency diagram (continuous)
6	a	Draw scatter graphs and understand correlation
	b	Construct, describe and interpret two way tables
	c	Construct and interpret network diagrams
7	a	Draw frequency polygons
	b	Construct and interpret flow charts with loops
	c	Draw lines of best fit on a scattergraph
8	a	Construct a c.f. curve and use to find median and I.Q. range

OAT 14

3	a	Place events in order of 'likelyhood'
	b	Understand evens and of outcomes being > or < likely than this
	c	Distinguish between fair and unfair
4	a	Understand and use the probability scale 0 to 1
	b	Give and justify estimates of probabilities
	c	List all possible outcomes of an event
5	a	Know different outcomes of an event are possible
	b	Distinguish between experimental & theoretical probability
	c	Know that n equally likely event have probability 1/n
6	a	List all possible outcomes of 2 combined events
	b	Know that sum of probabilities of all outcomes is 1 and p(A') = 1 - p(A)
7	a	Use relative frequency as an estimate of probability
	b	Recognise that subjective estimates of probability are sometimes required
	c	Use p(A u B) = p(A) + p(B)
8	a	Know that p(A n B) < p(A u B)
	b	Use tree diagrams to calculate combined probabilities

ledge, skills, standing and use nber, algebra and ures (ATs 9-14).

Attainment Target 12: Handling data

Pupils should collect, record and process data.

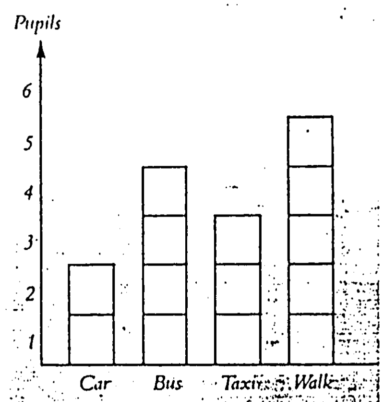
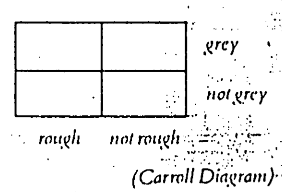
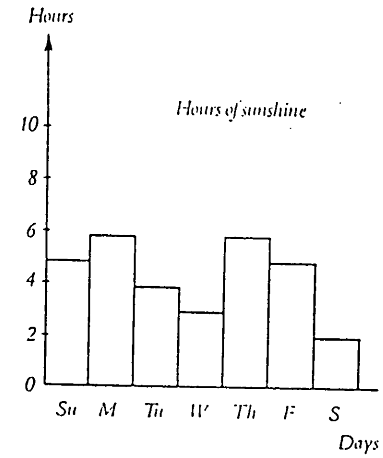
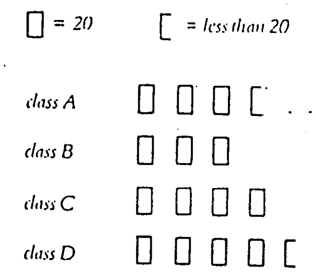
LEVEL	STATEMENTS OF ATTAINMENT	EXAMPLE
	Pupils should:	
1	• select criteria for sorting a set of objects and apply consistently.	
2	• choose criteria to sort and classify objects; record results of observations or outcomes of events.	*Identify those children who walk to school and those who travel by bus or car.*
	• help to design a data collection sheet and use it to record a set of data leading to a frequency table.	*Record the number and type of birds visiting the bird table:* *Blackbird XX 2* *Sparrow XXXXX 5* *Robin X 1* *Blue Tit XXX 3*
3	• extract specific pieces of information from tables and lists.	*Read off a value from a table, the cost of an item in a catalogue, etc.*
	• enter and access information in a simple database.	*Handle weather statistics or personal data, such as height, date of birth, age, etc.*
4	• specify an issue for which data are needed; collect, group and order discrete data using tallying methods with suitable equal class intervals and create a frequency table for grouped data.	*Find and record the number of pupils born in each month of the year.*
	• understand, calculate and use the mean and range of a set of data.	*Calculate the means to compare the scoring records of two hockey teams which have played different numbers of games.*
5	• design and use an observation sheet to collect data; collate and analyse results.	*Devise a simple habitat recorder for an ecological survey.* *Conduct a survey of cars passing with one, two, three, ... occupants.*
	• collect, group and order continuous data using equal class intervals and create a frequency table for grouped data.	*Collect information about height of children in a year group.* *Handle data arising through experiments or measurements in science, geography and CDT and from published sources in other areas of the curriculum.*
	• insert and interrogate data in a computer database and draw conclusions.	*Draw conclusions from census data about the effect of an epidemic.*
6	• specify an issue for which data are needed; design and use an appropriate observation sheet to collect data; collate and analyse results.	*Determine the best location for a pedestrian crossing.*
	• design and use a questionnaire to survey opinion (taking account of bias); collate and analyse results.	*Conduct a survey of taste in poetry, music, literature, art, television programmes, etc.*
7	• specify a simple hypothesis; design and use an appropriate questionnaire to test it (only yes/no responses required); collect and analyse the results to see whether the hypothesis is valid.	*Test the hypothesis that pupils/parents would prefer the school day to start at 0800 hours and finish at 1400 hours without a lunch break.*
	• using relevant data, record and organise grouped data into class intervals suitably defined; produce a frequency table; calculate the mean (using a calculator).	*Prepare tables; calculate the mean:* *(a) Measurement of heights:* *Use 10cm intervals from 120–200cm: class intervals defined as:* *Interval — Mid Interval Point* *$120 \leq h < 130$cm — 125cm* *$130 \leq h < 140$cm — 135cm* *$140 \leq h < 150$cm — 145cm, etc.* *h = height (centimetres)* *(b) Examination marks: Range 0–100; intervals of 10 marks:* *0- 9 mid point 4·5* *10-19 mid point 14·5* *20-29 mid point 24·5* *30-39 mid point 34·5*
	• find the mean, median, mode and range of frequency distribution for a given set of data and interpret the results.	*Compare the mean heights of sets of children of different ages and interpret.*

LEVEL	STATEMENTS OF ATTAINMENT	EXAMPLE
8	• design and use an appropriate questionnaire with three or more possible responses to each question; collate and analyse the results to test an hypothesis. • construct a cumulative frequency table.	*Analyse findings from a survey in which respondents have listed a set of products in order of preference.*
9	• use sampling to investigate a population and recognise the reliability of different sizes of samples.	
10	• describe the range of a variable using different measures of dispersion; calculate standard deviation of a set of data.	

...standing and use
...nber, algebra and
...ures (ATs 9-14).

Attainment Target 13: Handling data

Pupils should represent and interpret data.

LEVEL	STATEMENTS OF ATTAINMENT	EXAMPLE
	Pupils should:	
1	• record with real objects or drawings and comment about the result.	*Draw a simple picture to represent 'Children at our table'.*
	• create simple mapping diagrams showing relationships; read and interpret them.	*Link children's names to pets owned.*
2	• construct, read and interpret block graphs and frequency tables.	***Means of transport to school*** 
	• use diagrams to represent the result of classifying using two different criteria.	
3	• construct and interpret bar charts.	
	• create and interpret graphs (pictograms) where the symbol represents a group of units.	*Number of raffle tickets sold by each class* 

4

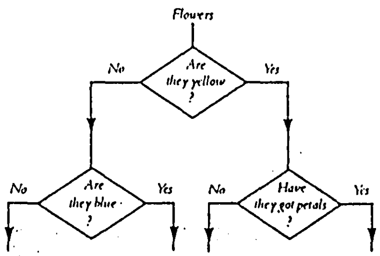
ᶠEL

STATEMENTS OF ATTAINMENT

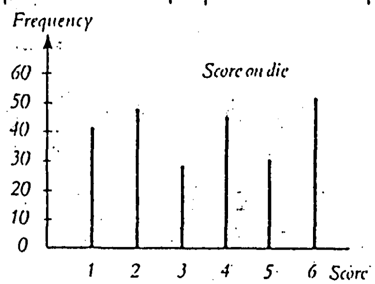
EXAMPLE

4

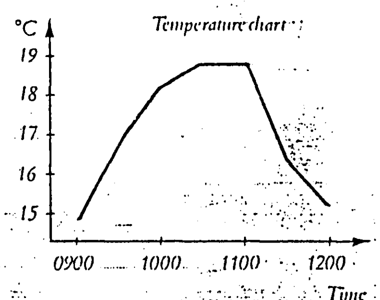
- create a decision tree-diagram with questions to sort and identify a collection of objects.



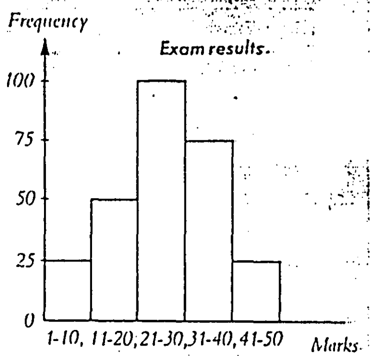
- construct, read and interpret a bar-line graph for a discrete variable (where the length of the bar-line represents the frequency).



- construct and interpret a line graph and know that the intermediate values may or may not have a meaning.

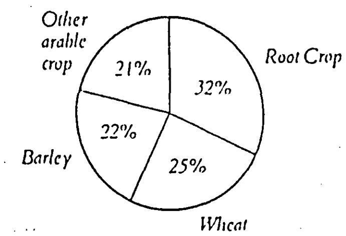


- construct and interpret a frequency diagram choosing suitable class intervals covering the range for a discrete variable.



5

- construct and interpret a pie chart from a collection of data with a few variables; interpret pie charts already presented in journals or newspapers.



Farming Statistics

- construct and interpret conversion graphs.

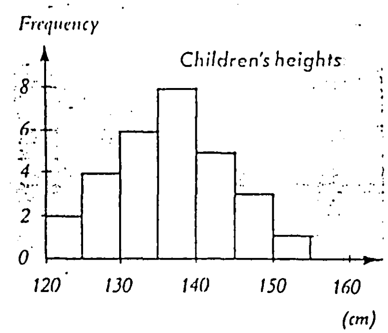
Devise and use conversion graphs for different currencies, and for converting thermometer readings from Fahrenheit to Celsius.

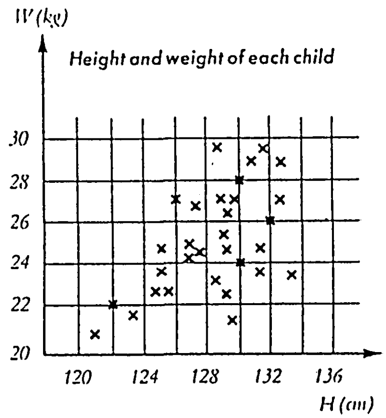
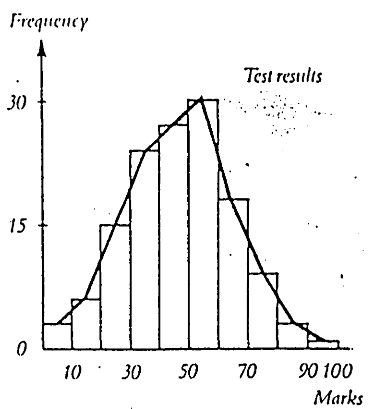
- construct and interpret a frequency diagram choosing suitable class intervals covering the range for a continuous variable.

Class intervals	*Frequency*
$120 \leq h < 125$	2
$125 \leq h < 130$	4
$130 \leq h < 135$	7
$135 \leq h < 140$	8
$140 \leq h < 145$	5
$145 \leq h < 150$	3
$150 \leq h < 155$	1

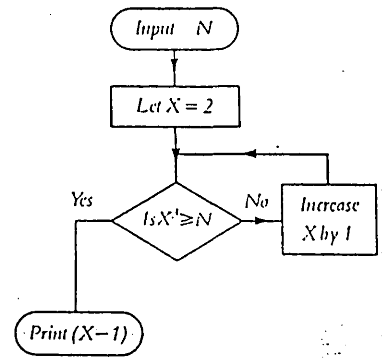
h = height (centimetres)

5



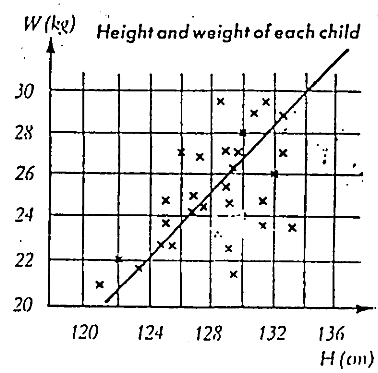
LEVEL	STATEMENTS OF ATTAINMENT	EXAMPLE
5	• create scatter graphs for discrete and continuous variables and have a basic understanding of correlation.	
	• construct, describe and interpret information through two-way tables.	*Record the orders from 10 people for various alternatives in a restaurant.*
	• construct and interpret network diagrams which represent relationships or connections.	*Use network diagrams to represent railway/airline connections between a number of major cities.*
7	• draw a frequency polygon as a line graph from a frequency distribution for grouped data; make comparisons between two frequency distributions.	

• construct and interpret flow diagrams with and without loops.

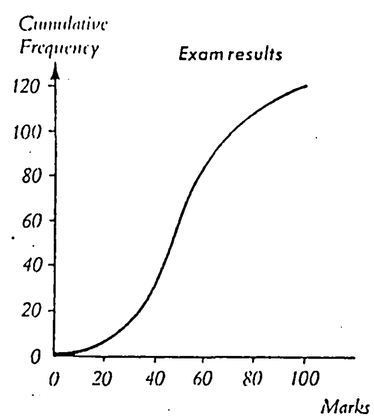


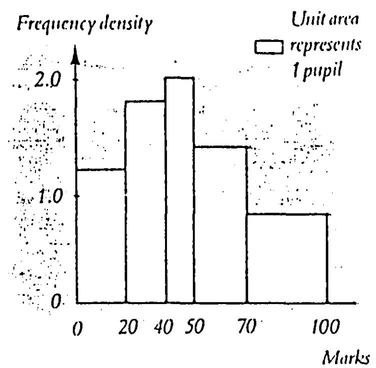
To find the largest whole number less than the cube root of a positive whole number N

• draw a line of 'best fit' by inspection on a scatter diagram.

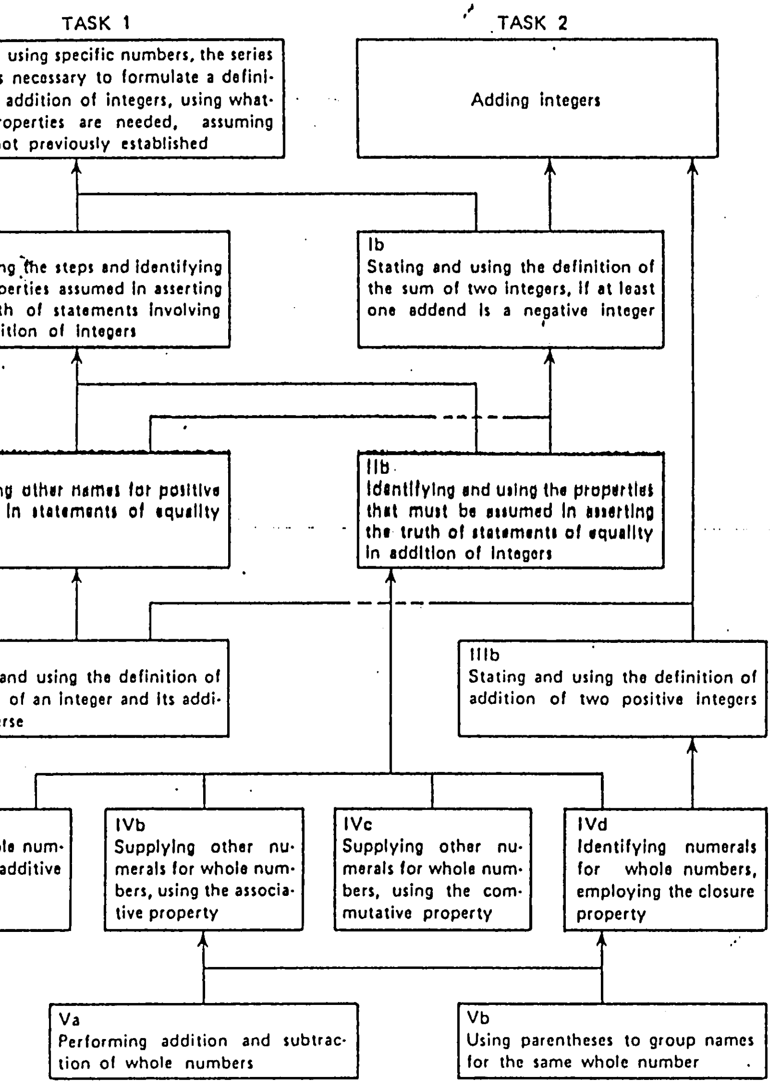


• construct a cumulative frequency curve using the upper boundary of the class interval; find the median, the upper quartile, the lower quartile and the interquartile range; interpret the results.



9	• present a set of complex data in a simplified form, using a variety of diagrams and graphs and computer statistical packages. • construct and interpret a histogram with understanding of the connection between area and frequency.	*Make a tal[illegible]om a government report or an opinion poll and prepare it for publication in the popular press.* *Distribution of marks in an exam* 
10	• interpret various types of diagrams such as those used in critical path analysis and linear programming. • consider different shapes of histograms representing distributions with special reference to mean and dispersion, including the normal distribution.	*Use critical path diagrams in preparing for a social event or a business venture.*

APPENDIX 2: Learning hierarchies as proposed by Gagné in illustration of the Cumulative Learning Theory



. 3.1 A learning hierarchy for two tasks related to adding integers. (From gné, Mayor, Garstens, & Paradise, 1962. Copyright 1962 by the American chological Association. Reprinted by permission.)

A

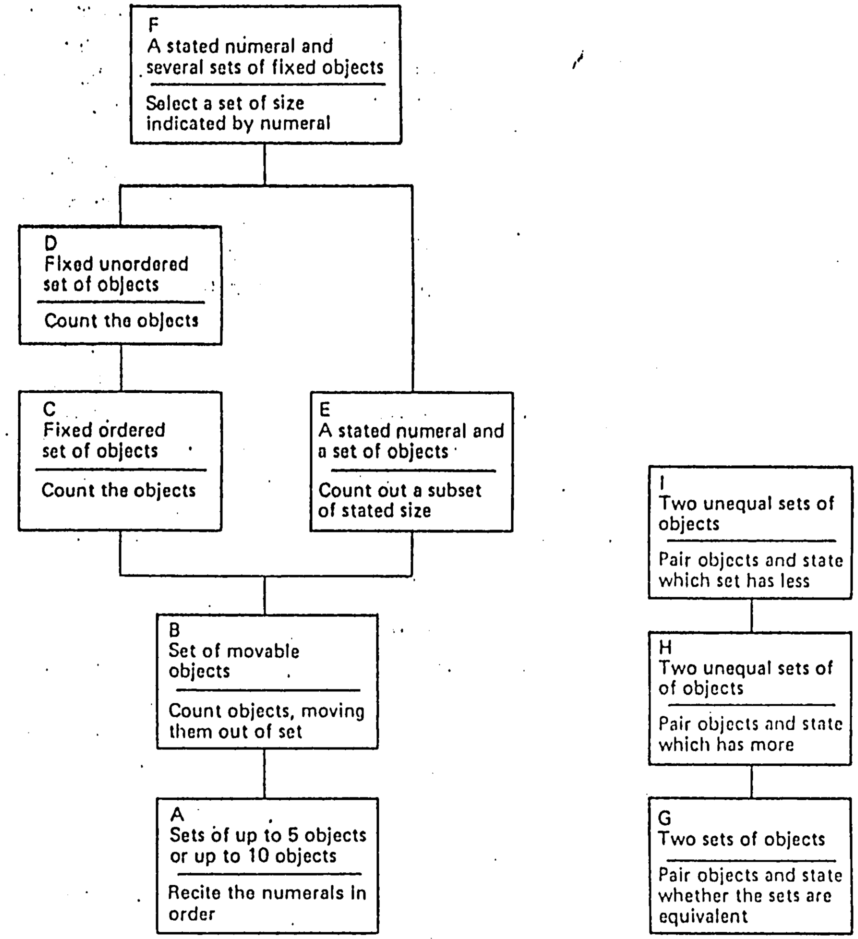


FIG. 3.4 Learning hierarchies for counting and one-to-one correspondence used in developing an early mathematics curriculum. (Adapted from Resnick et al., 1973.)

2B

9

APPENDIX 3: (i) Conversion chart of Havering Index raw scores to chronological reading age.

(ii) Analysis sheets for Havering Index for each of the five tests used in main testing.

TOTAL OF 4 COLUMNS

	0	1	2	3	4	5	6	7	8	9	9+
71			14.0	13.1	12.2	11.7	11.0	10.7	10.3	10.1	9.8
70			13.8	13.0	12.1	11.6	10.9	10.6	10.3	10.0	9.7
69			13.7	12.8	12.0	11.5	10.8	10.5	10.2	9.9	9.7
68			13.6	12.7	11.9	11.4	10.8	10.4	10.2	9.8	9.6
67			13.5	12.5	11.8	11.3	10.7	10.3	10.1	9.8	9.5
66		15.0	13.3	12.4	11.7	11.2	10.7	10.3	10.0	9.7	9.4
65		14.8	13.0	12.2	11.6	11.0	10.6	10.2	9.9	9.7	9.4
64		14.7	12.9	12.1	11.5	10.9	10.5	10.2	9.8	9.6	9.3
63		14.5	12.8	12.0	11.3	10.8	10.4	10.1	9.8	9.5	9.3
62		14.3	12.7	11.9	11.2	10.7	10.3	10.0	9.7	9.5	9.3
61		14.0	12.5	11.8	10.0	10.6	10.2	9.9	9.6	9.4	9.2
60		13.8	12.3	11.7	10.9	10.5	10.1	9.8	9.5	9.4	9.2
59	15.0	13.7	12.2	11.5	10.8	10.4	10.0	9.8	9.5	9.3	9.2
58	14.8	13.5	12.0	11.3	10.7	10.3	9.9	9.7	9.4	9.3	9.1
57	14.7	13.3	11.9	11.2	10.7	10.2	9.8	9.6	9.4	9.2	9.1
56	14.5	13.2	11.7	11.0	10.6	10.1	9.8	9.5	9.3	9.2	9.1
55	14.3	13.0	11.6	10.9	10.5	10.0	9.7	9.4	9.3	9.1	9.0
54	14.0	12.8	11.5	10.8	10.4	9.9	9.6	9.3	9.2	9.1	8.9
53	13.7	12.7	11.3	10.7	10.3	9.8	9.5	9.3	9.2	9.0	8.9
52	13.5	12.5	11.2	10.6	10.2	9.8	9.4	9.2	9.1	8.9	8.8
51	13.3	12.3	11.0	10.5	10.0	9.7	9.3	9.2	9.1	8.9	8.8
50	13.2	12.0	10.9	10.4	9.9	9.6	9.3	9.1	9.0	8.8	8.7
49	13.0	11.8	10.8	10.3	9.8	9.5	9.2	9.1	8.9	8.8	8.7
48	12.8	11.7	10.7	10.2	9.7	9.4	9.2	9.0	8.9	8.7	8.6
47	12.5	11.5	10.6	10.0	9.6	9.3	9.1	8.9	8.8	8.7	8.6
46	12.3	11.3	10.5	9.9	9.5	9.2	9.1	8.9	8.8	8.6	8.5
45	12.2	11.2	10.4	9.8	9.4	9.2	9.0	8.8	8.7	8.6	8.5
44	12.0	11.0	10.3	9.7	9.3	9.1	8.9	8.8	8.7	8.5	8.4
43	11.8	10.8	10.1	9.6	9.3	9.0	8.9	8.7	8.6	8.4	8.4
42	11.5	10.7	10.0	9.5	9.2	8.9	8.8	8.7	8.6	8.4	8.4
41	11.4	10.6	9.8	9.4	9.1	8.9	8.8	8.6	8.5	8.4	8.3
40	11.2	10.5	9.7	9.3	9.0	8.8	8.7	8.6	8.5	8.3	8.3
39	11.1	10.3	9.6	9.2	8.9	8.8	8.7	8.5	8.4	8.3	8.3
38	11.0	10.2	9.5	9.1	8.9	8.7	8.6	8.5	8.4	8.3	8.2
37	10.8	10.1	9.4	9.0	8.8	8.7	8.6	8.4	8.4	8.2	8.2
36	10.5	10.0	9.3	8.9	8.8	8.6	8.5	8.4	8.3	8.2	8.2
35	10.3	9.8	9.2	8.8	8.7	8.6	8.5	8.3	8.3	8.2	8.1
34	10.2	9.7	9.0	8.8	8.7	8.5	8.4	8.3	8.2	8.1	8.1
33	10.0	9.5	8.9	8.7	8.6	8.5	8.4	8.3	8.2	8.1	8.1
32	9.9	9.4	8.8	8.7	8.6	8.4	8.3	8.2	8.1	8.1	8.0
31	9.8	9.3	8.8	8.6	8.5	8.4	8.3	8.2	8.1	8.0	8.0
30	9.6	9.2	8.7	8.6	8.4	8.3	8.2	8.1	8.0	8.0	7.9
29	9.5	9.1	8.7	8.5	8.4	8.3	8.2	8.1	8.0	7.9	7.9
28	9.3	9.0	8.6	8.4	8.3	8.2	8.1	8.0	7.9	7.8	7.8
27	9.2	8.9	8.5	8.4	8.3	8.2	8.1	8.0	7.8	7.8	7.8
26	9.1	8.8	8.4	8.3	8.2	8.1	8.0	7.9	7.8	7.7	7.7
25	9.0	8.7	8.4	8.3	8.2	8.1	7.9	7.8	7.7	7.7	7.6
24	8.9	8.6	8.3	8.2	8.1	8.0	7.8	7.7	7.7	7.7	7.6
23	8.8	8.5	8.3	8.2	8.1	7.9	7.8	7.7	7.6	7.6	7.5
22	8.7	8.4	8.2	8.1	8.0	7.8	7.7	7.6	7.6	7.6	7.5
21	8.5	8.3	8.2	8.1	7.9	7.8	7.7	7.6	7.5	7.5	7.4
20	8.4	8.3	8.1	8.0	7.8	7.7	7.6	7.5	7.5	7.4	7.3
19	8.3	8.2	8.1	7.9	7.8	7.7	7.6	7.5	7.4	7.3	7.2
18	8.3	8.2	8.0	7.8	7.7	7.6	7.5	7.4	7.3	7.2	7.1
17	8.2	8.1	7.9	7.7	7.6	7.5	7.4	7.3	7.2	7.1	7.0
16	8.2	8.0	7.8	7.6	7.5	7.4	7.3	7.2	7.1	7.0	6.9
15	8.1	7.9	7.7	7.5	7.4	7.3	7.2	7.1	7.0	6.9	6.8
14	8.0	7.8	7.6	7.4	7.3	7.2	7.1	7.0	6.9	6.8	6.7
13	7.9	7.7	7.5	7.3	7.2	7.1	7.0	6.9	6.8	6.7	6.6
12	7.8	7.6	7.4	7.2	7.1	7.0	.9	6.8	6.7	6.6	6.5
11	7.7	7.5	7.3	7.1	7.0	6.9	6.8	6.7	6.6	6.5	6.4
10	7.6	7.4	7.2	7.0	6.9	6.8	6.7	6.6	6.5	6.4	6.3
9	7.5	7.3	7.1	6.9	6.8	6.7	6.6	6.5	6.4	6.3	6.2
8	7.4	7.2	7.0	6.8	6.7	6.6	6.5	6.4	6.3	6.2	6.1
7	7.3	7.1	6.9	6.7	6.6	6.5	6.4	6.3	6.2	6.1	6.0
6	7.2	7.0	6.8	6.6	6.5	6.4	6.3	6.2	6.1	6.0	5.9
5	7.1	6.9	6.7	6.5	6.4	6.3	6.2	6.1	6.0	5.9	5.8
4	7.0	6.8	6.6	6.4	6.3	6.2	6.1	6.0	5.9	5.8	5.7
3	6.9	6.7	6.5	6.3	6.2	6.1	6.0	5.9	5.8	5.7	5.6
2	6.8	6.6	6.4	6.2	6.1	6.0	5.9	5.8	5.7	5.6	5.5
1	6.7	6.5	6.3	6.1	6.0	5.9	5.8	5.7	5.6	5.5	5.4

Number of sentences

Total of 4 columns

	0	1	2	3	4	5	6	7	8	9	9+
											15.0
135											14.9
134											14.
133											14.[illegible]
132											14.7
131											14.7
130											14.6
129											14.5
128										15.0	14.4
127										14.9	14.3
126										14.8	14.3
125										14.7	14.2
124										14.7	14.1
123										14.6	14.0
122										14.5	13.9
121									15.0	14.4	13.8
120									14.9	14.3	13.6
119									14.8	14.2	13.5
118									14.8	14.1	13.4
117									14.7	14.0	13.3
116									14.6	13.9	13.2
115									14.5	13.8	13.2
114									14.4	13.6	13.1
113								15.0	14.3	13.5	13.0
112								14.9	14.2	13.4	12.9
111								14.8	14.1	13.3	12.8
110								14.7	14.0	13.2	12.7
109								14.6	13.9	13.2	12.6
108								14.5	13.8	13.1	12.5
107								14.4	13.7	13.0	12.4
106							15.0	14.2	13.5	12.9	12.3
105							14.9	14.1	13.4	12.8	12.3
104							14.8	14.0	13.3	12.7	12.2
103							14.7	13.9	13.2	12.6	12.1
102							14.6	13.8	13.1	12.5	12.0
101							14.5	13.6	13.0	12.4	11.0
100							14.4	13.5	12.9	12.3	11.8
99							14.3	13.4	12.8	12.3	11.8
98						15.0	14.1	13.3	12.7	12.2	11.7
97						14.9	14.0	13.2	12.6	12.1	11.7
96						14.8	13.9	13.1	12.5	12.0	11.6
95						14.7	13.8	13.0	12.4	11.9	11.6
94						14.5	13.7	12.9	12.3	11.8	11.5
93						14.3	13.5	12.8	12.2	11.8	11.4
92						14.2	13.3	12.7	12.1	11.7	11.3
91						14.1	13.2	12.6	12.0	11.7	11.2
90					15.0	14.0	13.1	12.5	11.9	11.6	11.2
89					14.8	13.9	13.0	12.4	11.8	11.5	11.1
88					14.7	13.7	12.9	12.3	11.8	11.4	11.0
87					14.5	13.6	12.8	12.2	11.7	11.3	10.9
86					14.3	13.5	12.7	12.1	11.7	11.2	10.8
85					14.2	13.3	12.6	12.0	11.6	11.1	10.8
84				15.0	14.0	13.2	12.5	11.9	11.5	11.0	10.7
83				14.9	13.8	13.0	12.4	11.8	11.4	10.9	10.6
82				14.8	13.7	12.9	12.2	11.7	11.3	10.8	10.5
81				14.7	13.6	12.8	12.1	11.7	11.2	10.8	10.4
80				14.5	13.5	12.7	12.0	11.6	11.1	10.7	10.4
79				14.3	13.3	12.6	11.9	11.5	11.0	10.6	10.3
78				14.2	13.2	12.5	11.8	11.4	10.9	10.5	10.3
77				14.0	13.1	12.4	11.7	11.3	10.8	10.4	10.2
76			15.0	13.8	13.0	12.3	11.6	11.2	10.7	10.3	10.2
75			14.8	13.7	12.8	12.1	11.5	11.0	10.7	10.3	10.1
74			14.7	13.5	12.7	12.0	11.4	10.9	10.6	10.2	10.0
73			14.5	13.3	12.5	11.9	11.3	10.8	10.5	10.2	9.9
72			14.3	13.2	12.4	11.8	11.1	10.8	10.4	10.1	9.8

HAVERING INDEX CHECKING SHEET

NAME OF BOOK Sorting & Grouping Skills

READING AGE 10.0

WORDS OF 8 OR MORE LETTERS	WORDS OF 7 LETTERS	WORDS OF 6 LETTERS	WORDS OF 5 LETTERS
4 SCIENTIST	2 PICTURE	2 AMAZON	1 SHOWS
8 RAINFOREST	5 SHOWING	3 NUMBER	2 FOUND
12 COMPLETE	7 DAISIES	5 COMMON	3 TABLE
16 SPOTTERS	10 GROWING	6 MARKED	4 SPOTS
20 " "	12 SQUARES	8 SQUARE	5 THREE
24 " "	15 DAISIES	9	6 THERE
28 BOTANIST	17 TEACHER	11	7 WHICH
32	20	12	8 THREE
36	22	14	9 SHOWS
40	25	15	10 FIELD
44	27	17	11 COUNT
48	30	18	12 TABLE
52	32	20	13 CLASS
56	35	21	14 GIVE
60	37	23	15 " "
64	40	24	16 WHICH
68	42	26	17 MARKS
72	45	27	18
76	47	29	19
80	50	30	20

NUMBER OF SENTENCES 9 TOTAL OF 4 COLUMNS 70

Select 100 words in centre of the book. Discount Personal Pronouns and Speech Areas.

Total end number of each column. Count sentences.

HAVERING INDEX CHECKING SHEET

NAME OF BOOK Measures of Location

READING AGE 10·5

WORDS OF 8 OR MORE LETTERS	WORDS OF 7 LETTERS	WORDS OF 6 LETTERS	WORDS OF 5 LETTERS
4 children	2 classes	2 weight	1 class
8 children	5 average	3 median	2 other
12 following	7 average	5 middle	3 class
16 questions	10 involve	6	4 which
20 balances	12 picture	8	5 class
24 triangle	15 balance	9	6 think
28 diagrams	17 balance	11	7 pivot
32 diagrams	20	12	8 shown
36 schoolchildren	22	14	9 these
40 blackboard	25	15	10 pivot
44	27	17	11 level
48	30	18	12 these
52	32	20	13 level
56	35	21	14 group
60	37	23	15 asked
64	40	24	16 write
68	42	26	17 their
72	45	27	18 sizes
76	47	29	19 wrote
80	50	30	20 group.

NUMBER OF SENTENCES 10 TOTAL OF 4 COLUMNS 82

Select 100 words in centre of the book. Discount Personal Pronouns and Speech Areas. Total end number of each column. Count sentences.

HAVERING INDEX CHECKING SHEET

NAME OF BOOK DISPERSION TEST

READING AGE 10·1

WORDS OF 8 OR MORE LETTERS	WORDS OF 7 LETTERS	WORDS OF 6 LETTERS	WORDS OF 5 LETTERS
4 children	2 picture	2 looked	1 grown
8 sunflowers	5 exactly	3 seemed	2 first
12 difference	7 picture	5 waists	3 shows
16 difference	10 heights	6 listed	4 after
20 children	12 someone	8	5 there
24 measured	15 measure	9	6 grown
28 commented	17 missing.	11	7 weeks
32 measurements	20	12	8 grown
36 discovered	22	14	9 shown
40	25	15	10 there
44	27	17	11 their
48	30	18	12 agree
52	32	20	13 group
56	35	21	14 their
60	37	23	15 range
64	40	24	16 affect
68	42	26	17
72	45	27	18
76	47	29	19
80	50	30	20

NUMBER OF SENTENCES 11 TOTAL OF 4 COLUMNS 75

Select 100 words in centre of the book. Discount Personal Pronouns and Speech Areas.
Total and number of each column. Count sentences.

HAVERING INDEX CHECKING SHEET

NAME OF BOOK: Bivariate Data

READING AGE: 9.2

WORDS OF 8 OR MORE LETTERS	WORDS OF 7 LETTERS	WORDS OF 6 LETTERS	WORDS OF 5 LETTERS
4 scientist	2 picture	2 medium	1 breed
8 examining	5 printed	3 either	2 moths
12 everyone	7 wearing	5 bigger	3 moths
16 newspaper	10	6 before	4 breed
20	12	8 likely	5 there
24	15	9	6 three
28	17	11	7 sizes
32	20	12	8 small
36	22	14	9 large
40	25	15	10 spots
44	27	17	11 think
48	30	18	12 moths
52	32	20	13 spots
56	35	21	14 other
60	37	23	15 round
64	40	24	16 write
68	42	26	17 there
72	45	27	18 table
76	47	29	19 below
80	50	30	20 years

about
shows
think
table
belts

NUMBER OF SENTENCES 9 TOTAL OF 4 COLUMNS 56

Select 100 words in centre of the book. Discount Personal Pronouns and Speech Areas. Total and number of each column. Count sentences.

HAVERING INDEX CHECKING SHEET

NAME OF BOOK: SAMPLING SKILLS

READING AGE: 10·6

WORDS OF 8 OR MORE LETTERS	WORDS OF 7 LETTERS	WORDS OF 6 LETTERS	WORDS OF 5 LETTERS	
4 collector	2 notices	2 insect	1 moths	
8 spotteng	5 further	3 select	2 which	
12 toadstool	7 balance	5 number	3 spots	
16 collects	10	6 choose	4 wants	
20 research	12	8 expert.	5 moths	
24 toadstools	15	9 spotty	6 which	
28 " "	17	11 spotty	7 spots	
32 explorer	20	12 circle	8 these	
36 discovery	22	14 choose	9 moths	
40	25	15 Amazon	10 there	
44	27	17 beetle	11 shade	
48	30	18	12 would	
52	32	20	13 types	
56	35	21	14 other	
60	37	23	15 plain	
64	40	24	16 wants	
68	42	26	17 which	
72	45	27	18 plain	which
76	47	29	19 whole	would
80	50	30	20 round	those

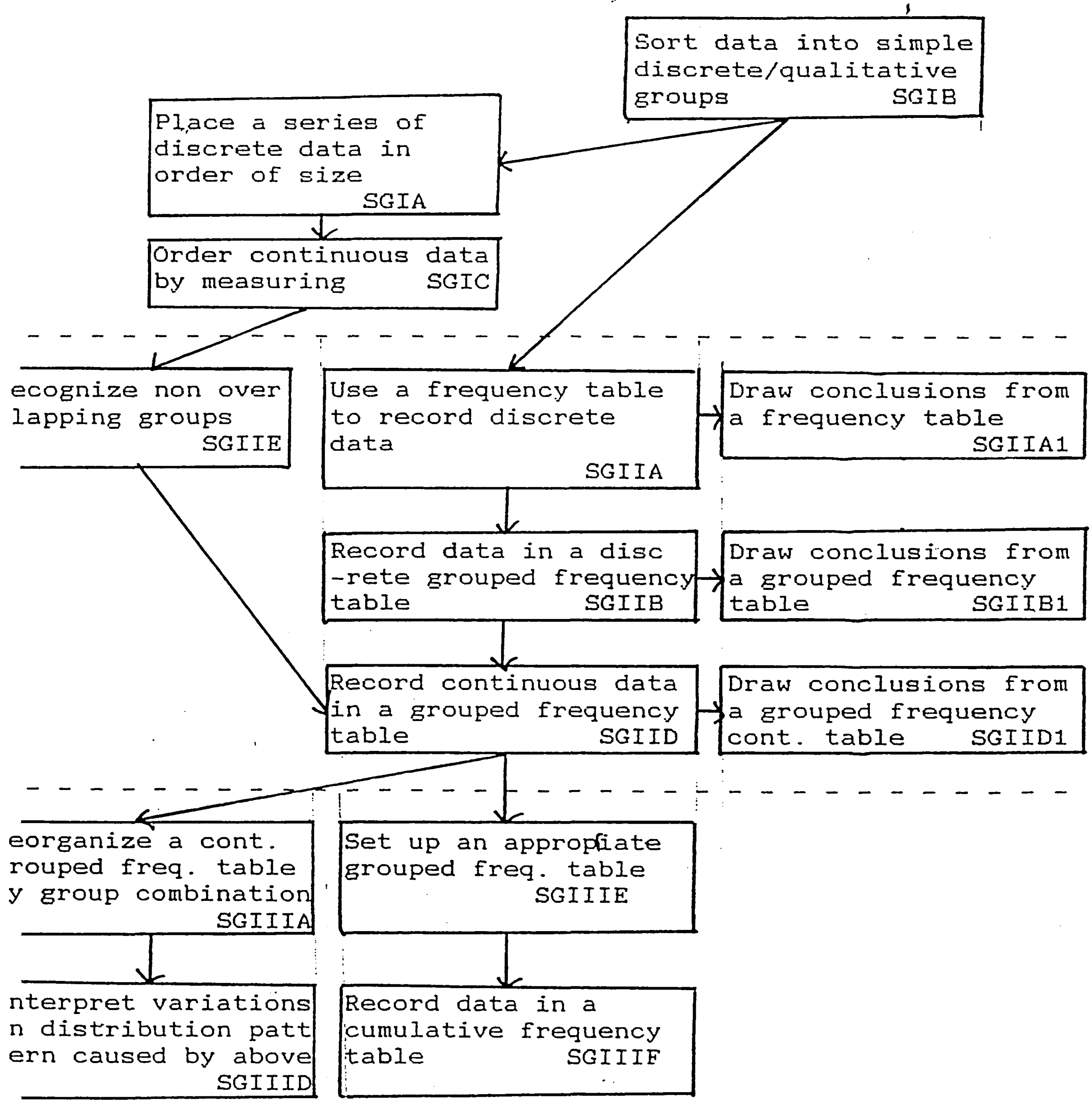
NUMBER OF SENTENCES 10 TOTAL OF 4 COLUMNS 83

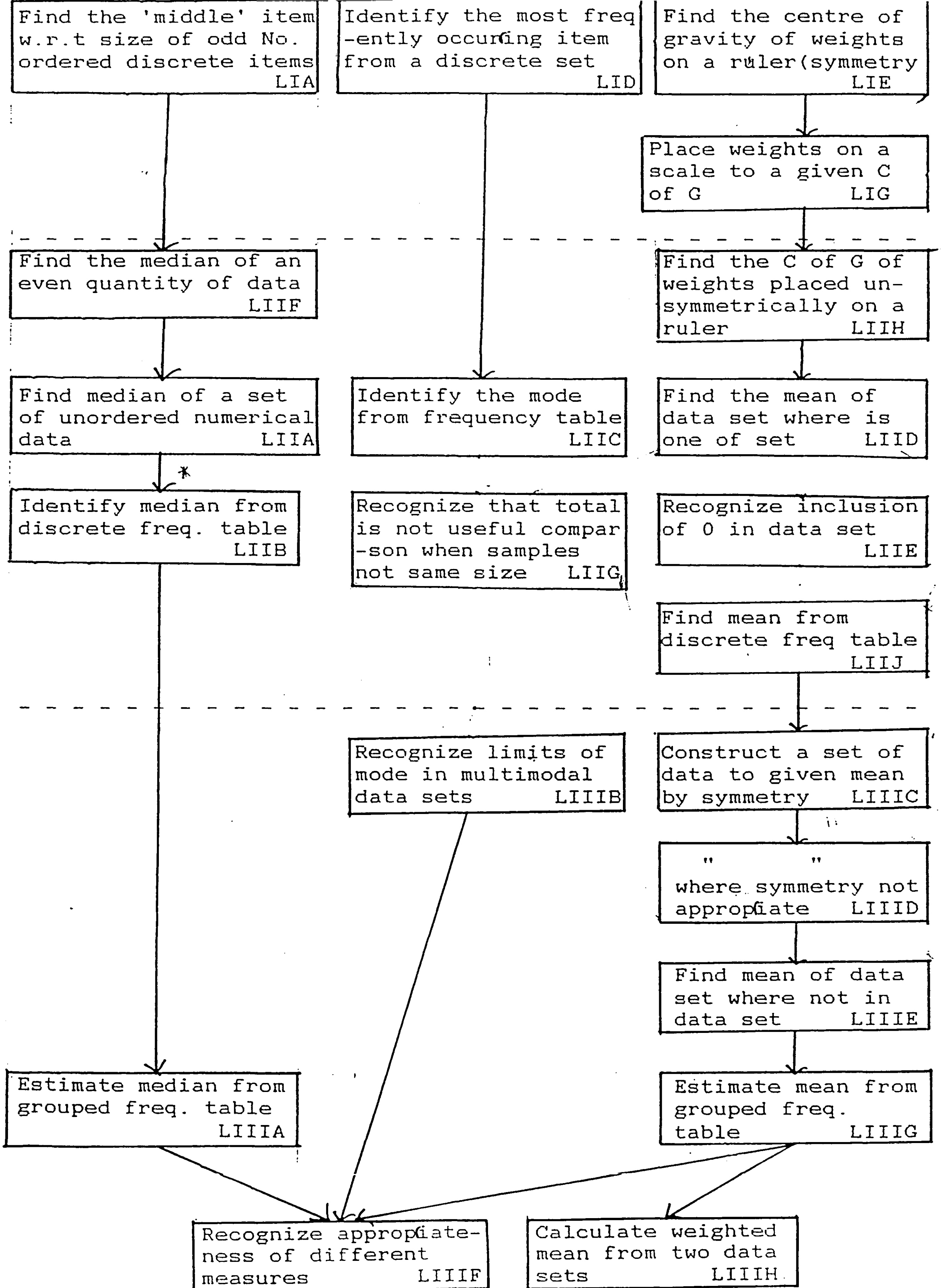
Select 100 words in centre of the book. Discount Personal Pronouns and Speech Areas.

Total end number of each column. Count sentences.

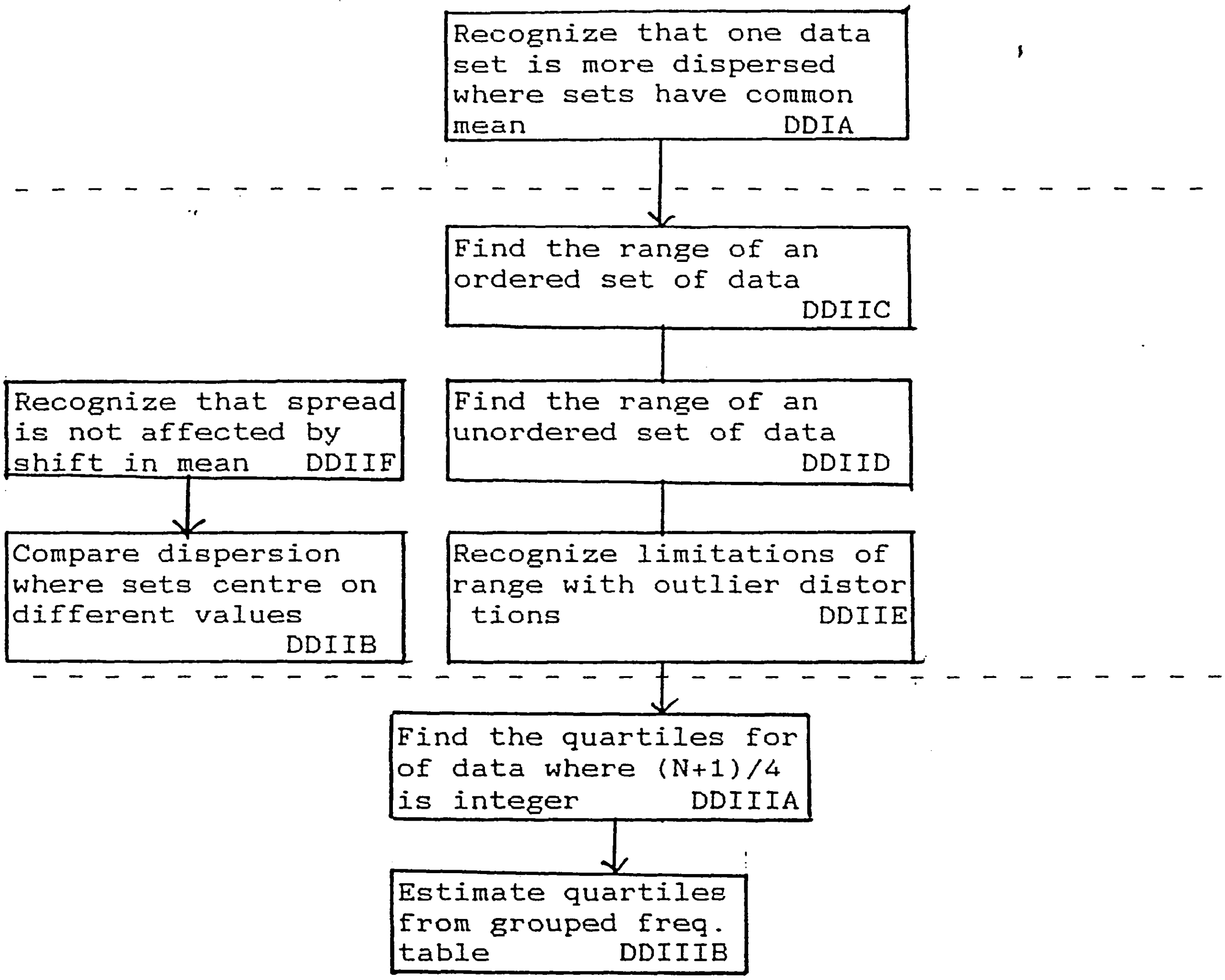
APPENDIX 4: Amended learning hierarchies for each of the five skill areas as used in the main testing.

Sorting & Grouping Skills

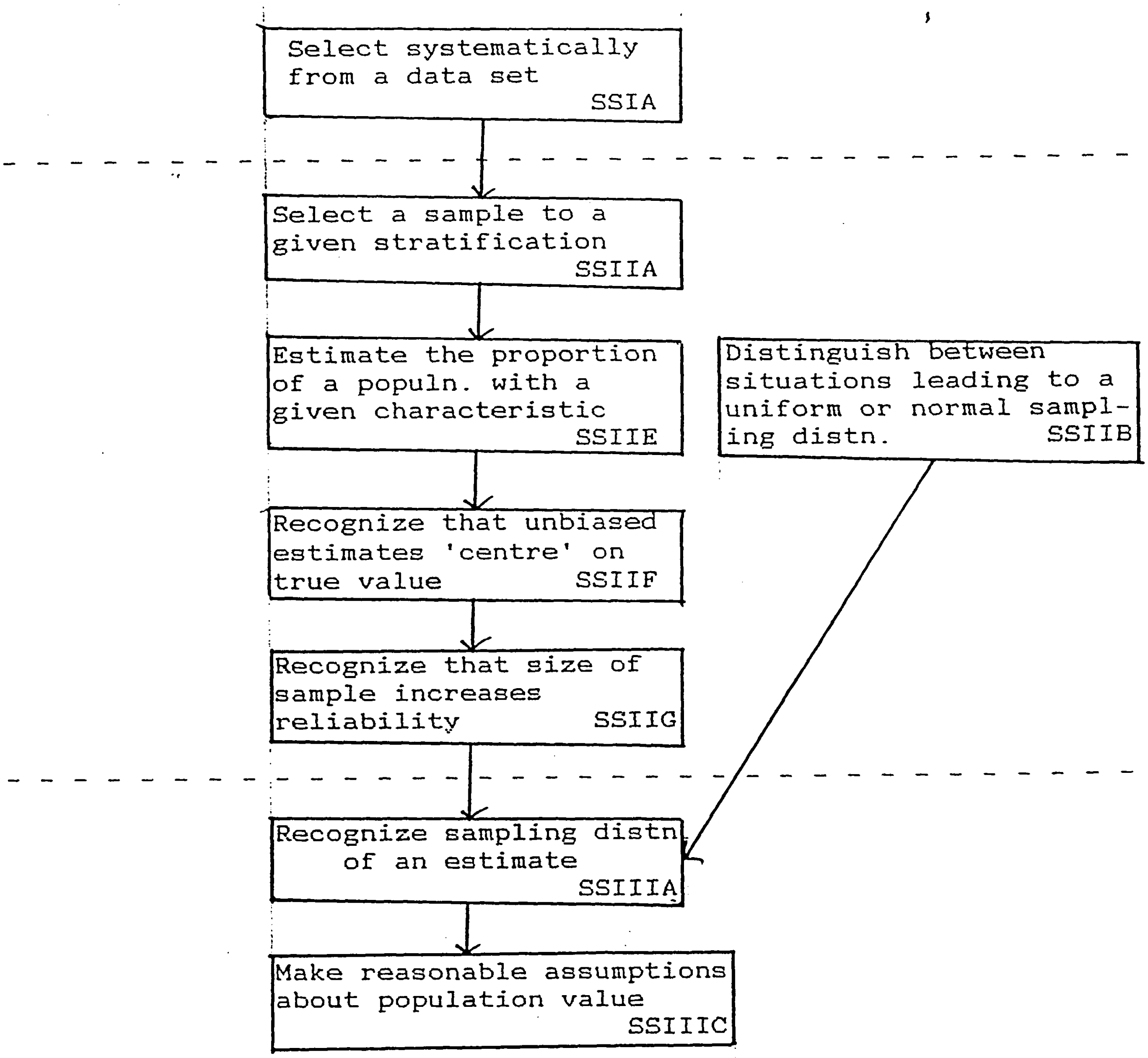



Find the 'middle' item w.r.t size of odd No. ordered discrete items LIA
Identify the most freq -ently occuring item from a discrete set LID
Find the centre of gravity of weights on a ruler(symmetry LIE
Place weights on a scale to a given C of G LIG
Find the median of an even quantity of data LIIF
Find the C of G of weights placed un-symmetrically on a ruler LIIH
Find median of a set of unordered numerical data LIIA
Identify the mode from frequency table LIIC
Find the mean of data set where is one of set LIID
Identify median from discrete freq. table LIIB
Recognize that total is not useful compar -son when samples not same size LIIG
Recognize inclusion of 0 in data set LIIE
Find mean from discrete freq table LIIJ
Recognize limits of mode in multimodal data sets LIIIB
Construct a set of data to given mean by symmetry LIIIC
" " where symmetry not appropiate LIIID
Find mean of data set where not in data set LIIIE
Estimate median from grouped freq. table LIIIA
Estimate mean from grouped freq. table LIIIG
Recognize appropiate-ness of different measures LIIIF
Calculate weighted mean from two data sets LIIIH

(I)-Distribution & Dispersion



Sampling Skills



APPENDIX 5: MINITAB macro written and used to carry out the White & Clarke Inclusion Analysis on data from the final testing.

MINITAB MACRO FOR CLARKE & WHITE INCLUSION TEST

This macro is operated by "execute 'clarke' "

First put a b c d e f in C1 - C6 where these are totals

```
                c
                b
                a
        f  e  d
```

Then let k1=N (total frequencies)

LISTING

```
let c7=2*c1/(2*c1+c2)
let c8=c5/(c5+2*c6)
let c9=(2*c1+c2)**2/(4*c1*k1)
let c10=1-(c5+2*c6)**2/(4*c6*k1)
let c11=c9-c10
let c12=1-c9
let c13=c12*c8**2+(c11*(1-c7)**2*c8**2)
let c14=(1-c13)**k1
let c15=c14+k1*(1-c13)**(k1-1)*c13
let c16=c15+(k1*(k1-1)/2)*(1-c13)**(k1-2)*c13**2
let c17=c16+(k1*(k1-1)*(k1-2))/6*(1-c13)**(k1-3)*c13**3
name c14 'f0'
name c15 'f<1'
name c16 'f<2'
name c17 'f<3'
paper
print c14-c17
nopaper
end
```

Results indicate the significance.

APPENDIX 6: Detailed Inclusion Analysis results as carried out on data from the final testing and used to produce the final hierarchies.

(i) Tables show number of candidates scoring 0,1,2 in skills being matched

(ii)Table shows significance level of various critical values in the f_{02} cell.

	0	1	2	ALL
	3	2	13	18
	0	0	2	2
	0	5	28	33
	3	7	43	53

6

	0	1	2	ALL
0	1	1	0	2
1	1	2	3	6
2	1	4	40	45
ALL	3	7	43	53

: SGIIE COLUMNS: SGIIIA

	0	1	2	ALL
	5	10	3	18
	0	2	0	2
	2	10	21	33
	7	22	24	53

ROWS: SGIIIE COLUMNS: SGIIIF

7

	0	2	ALL
0	23	2	25
1	2	1	3
2	11	14	25
ALL	36	17	53

SGIIE COLUMNS: SGIIIE

	0	1	2	ALL
	14	0	4	18
	2	0	0	2
	9	3	21	33
	25	3	25	53

ROWS: SGIIIA COLUMNS: SGIIID

8

	0	1	2	ALL
0	7	0	0	7
1	18	2	2	22
2	9	1	14	24
ALL	34	3	16	53

: SGIIA COLUMNS: A1

	0	1	2	ALL
	0	0	2	2
	1	0	5	6
	0	2	43	45
	1	2	50	53

ROWS: SGIID COLUMNS: D1

9

	0	1	2	ALL
0	3	0	0	3
1	4	3	0	7
2	0	12	31	43
ALL	7	15	31	53

: SGIIB COLUMNS: B1

	0	1	2	ALL
	0	1	1	2
	0	0	6	6
	2	1	42	45
	2	2	49	53

ROW	f0	f<1	f<2	f<3
1	0.00416	0.02816	0.09620	0.22227
2	0.00076	0.00658	0.02858	0.08289
3	0.94400	0.99843	0.99997	1.00000
4	0.63503	0.92462	0.98941	0.99889
5	0.81746	0.98254	0.99889	0.99995
6	0.58360	0.89950	0.98338	0.99794
7	1.00000	1.00000	1.00000	1.00000
8	0.99036	0.99995	1.00000	1.00000
9	0.46143	0.82093	0.95833	0.99266

	0	1	2	ALL
	11	0	1	12
	8	3	3	14
	11	7	6	24
	30	10	10	50

6

	0	1	2	ALL
0	24	1	0	25
1	8	1	1	10
2	2	0	13	15
ALL	34	2	14	50

: LIIC COLUMNS: LIIG

	0	1	2	ALL
	2	2	1	5
	5	9	3	17
	4	5	19	28
	11	16	23	50

ROWS: LIIE COLUMNS: LIIIE

7

	0	1	2	ALL
0	23	4	2	29
1	0	1	0	1
2	5	5	10	20
ALL	28	10	12	50

: LIE COLUMNS: LIG

	0	1	2	ALL
	4	0	1	5
	0	1	0	1
	2	24	18	44
	6	25	19	50

ROWS: LIIIC COLUMNS: LIIID

8

	0	1	2	ALL
0	24	1	0	25
1	8	1	1	10
2	2	0	13	15
ALL	34	2	14	50

: LIIH COLUMNS: LIID

	0	1	2	ALL
	6	5	1	12
	4	4	6	14
	5	6	13	24
	15	15	20	50

ROWS: LIID COLUMNS: LIIIE

9

	0	1	2	ALL
0	12	3	0	15
1	11	1	3	15
2	5	6	9	20
ALL	28	10	12	50

S: LIIG COLUMNS: LIIIB

	0	2	ALL
	6	5	11
	12	4	16
	19	4	23
	37	13	50

ROWS: LIIG COLUMNS: LIID

10

	0	1	2	ALL
0	7	3	1	11
1	3	5	8	16
2	5	7	11	23
ALL	15	15	20	50

ROW	f0	f<1	f<2	f<3
1	0.789849	0.976626	0.99827	0.99991
2	0.483607	0.837504	0.96440	0.99412
3	0.096630	0.327803	0.59879	0.80625
4	0.284524	0.646687	0.87257	0.96458
5	0.000123	0.001336	0.00720	0.02571
6	0.980493	0.999811	1.00000	1.00000
7	0.511675	0.856839	0.97093	0.99556
8	0.980493	0.999811	1.00000	1.00000
9	0.721480	0.957777	0.99570	0.99967
10	0.328144	0.697901	0.90206	0.97567

WS: DDIA COLUMNS: DDIIC

	0	1	2	ALL
0	7	1	2	10
1	7	1	3	11
2	7	1	20	28
L	21	3	25	49

OWS: DDIIC COLUMNS: DDIID

	0	1	2	ALL
0	21	0	0	21
1	0	3	0	3
2	1	3	21	25
LL	22	6	21	49

WS: DDIID COLUMNS: DDIIE

	0	1	2	ALL
0	6	9	7	22
1	4	2	0	6
2	3	9	9	21
L	13	20	16	49

OWS: DDIIE COLUMNS: DDIIIA

	0	1	2	ALL
0	12	1	0	13
1	10	5	5	20
2	13	0	3	16
LL	35	6	8	49

OWS: DDIIF COLUMNS: DDIIB

	0	1	2	ALL
0	14	0	2	16
1	7	1	10	18
2	0	3	12	15
LL	21	4	24	49

OWS: DDIIC COLUMNS: DDIIF

	0	1	2	ALL
0	9	9	3	21
1	1	2	0	3
2	6	7	12	25
LL	16	18	15	49

ROW	f0	f<1	f<2	f<3
1	0.959370	0.999181	0.99999	1.00000
2	0.739057	0.963224	0.99653	0.99976
3	0.013311	0.073415	0.20634	0.39823
4	0.928726	0.997449	0.99994	1.00000
5	0.908683	0.995783	0.99987	1.00000
6	0.061070	0.236770	0.48436	0.71211

S: IA COLUMNS: IIB

	1	2	ALL
	4	4	8
	14	32	46
	18	36	54

ROWS: IIL COLUMNS: IIIB

6	0	1	2	ALL
0	24	6	2	32
1	7	3	6	16
2	2	3	1	6
ALL	33	12	9	54

IIB COLUMNS: IIC

	0	1	2	ALL
	4	14	0	18
	0	4	32	36
	4	18	32	54

ROWS: IIIB COLUMNS: IIID

7	0	1	2	ALL
0	19	5	9	33
1	2	1	9	12
2	0	2	7	9
ALL	21	8	25	54

: IIA COLUMNS: IIJ

	0	1	ALL
	4	1	5
	4	2	6
	23	20	43
	31	23	54

ROWS: IIF COLUMNS: IIIE

8	0	1	2	ALL
0	22	11	2	35
1	2	4	0	6
2	2	4	7	13
ALL	26	19	9	54

IIA COLUMNS: IIIF

	0	1	2	ALL
	3	1	1	5
	3	1	2	6
	8	9	26	43
	14	11	29	54

ROWS: IB COLUMNS: IIIK

9	0	1	2	ALL
0	12	0	0	12
1	6	10	1	17
2	3	8	14	25
ALL	21	18	15	54

IIF COLUMNS: IIL

	0	1	2	ALL
	27	7	1	35
	3	3	0	6
	2	6	5	13
	32	16	6	54

ROW	f0	f<1	f<2	f<3
1	*	*	*	*
2	1.25207	0.971194	1.00211	0.999888
3	0.69165	0.947517	0.99397	0.999483
4	0.67700	0.942041	0.99296	0.999362
5	0.24349	0.591998	0.83680	0.949267
6	0.49619	0.846183	0.96733	0.994765
7	0.47203	0.828863	0.96124	0.993363
8	0.07793	0.281589	0.54275	0.761810
9	0.35876	0.730037	0.91860	0.981235

(1)

0	1	2	ALL
3	4	3	10
2	6	8	16
0	2	19	21
5	12	30	47

ROWS: IIG COLUMNS: IIIA

(4)	0	1	2	ALL
0	15	3	1	19
1	10	1	3	14
2	7	2	5	14
ALL	32	6	9	47

S: IIE COLUMNS: IIF

(2) 0	1	2	ALL
3	2	0	5
1	9	2	12
4	12	14	30
8	23	16	47

ROWS: IIIA COLUMNS: IIIC

(5)	0	1	2	ALL
0	23	9	0	32
1	4	1	1	6
2	1	2	6	9
ALL	28	12	7	47

: IIF COLUMNS: IIG

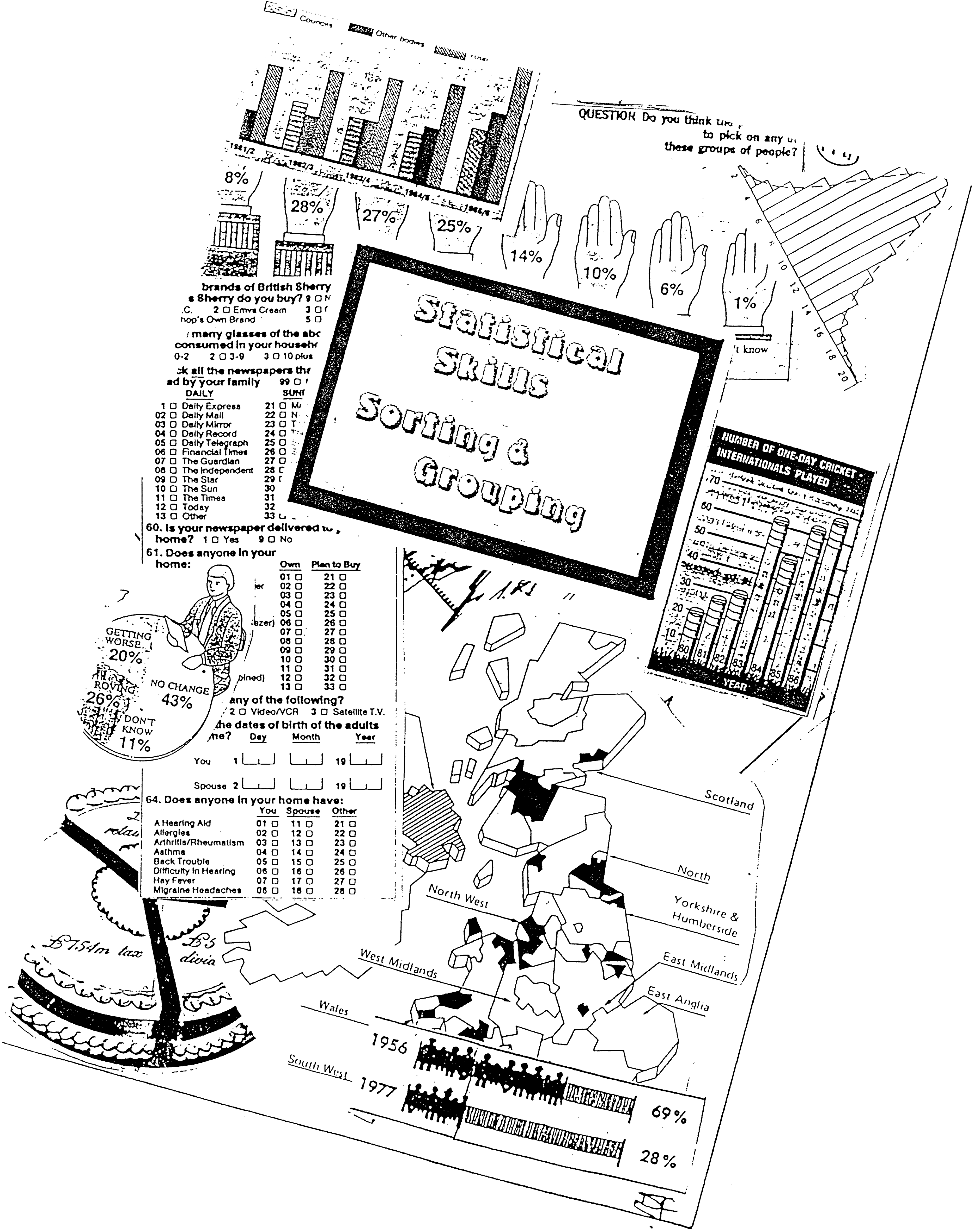
0	1	2	ALL
6	1	1	8
12	6	5	23
1	7	8	16
19	14	14	47

ROWS: IIB COLUMNS: IIIA

(6)	0	1	2	ALL
0	18	3	1	22
1	8	3	2	13
2	6	0	6	12
ALL	32	6	9	47

ROW	f0	f<1	f<2	f<3
1	0.080007	0.287603	0.551202	0.76949
2	0.169093	0.475378	0.746868	0.90381
3	0.645883	0.929538	0.990500	0.99904
4	0.875762	0.992106	0.999670	0.99999
5	0.369393	0.741194	0.924326	0.98315
6	0.857292	0.989511	0.999491	0.99998

APPENDIX 7: Tests used in the main testing after modification by pilot testing and expert analysis



Statistical Skills
Sorting & Grouping
8%
28%
27%
25%
14%
10%
6%
1%
QUESTION Do you think
to pick on any
these groups of people?
NUMBER OF ONE-DAY CRICKET INTERNATIONALS PLAYED
YEAR
60. Is your newspaper delivered to your home? 1 Yes 9 No
61. Does anyone in your home:
Own
Plan to Buy
64. Does anyone in your home have:
You
Spouse
Other
A Hearing Aid
Allergies
Arthritis/Rheumatism
Asthma
Back Trouble
Difficulty in Hearing
Hay Fever
Migraine Headaches
DAILY
Daily Express
Daily Mail
Daily Mirror
Daily Record
Daily Telegraph
Financial Times
The Guardian
The Independent
The Star
The Sun
The Times
Today
Other
GETTING WORSE 20%
NO CHANGE 43%
DON'T KNOW 11%
26%
£754m tax
Scotland
North
North West
Yorkshire & Humberside
West Midlands
East Midlands
East Anglia
Wales
South West
1956
1977
69%
28%

ANSWER AS MANY QUESTIONS AS YOU CAN - DO NOT WORRY IF THERE ARE SOME QUESTIONS YOU CANNOT ANSWER.

THIS TEST IS DESIGNED TO FIND OUT WHAT PEOPLE OF YOUR AGE KNOW

TAKE CARE WITH QUESTIONS YOU FIND EASY

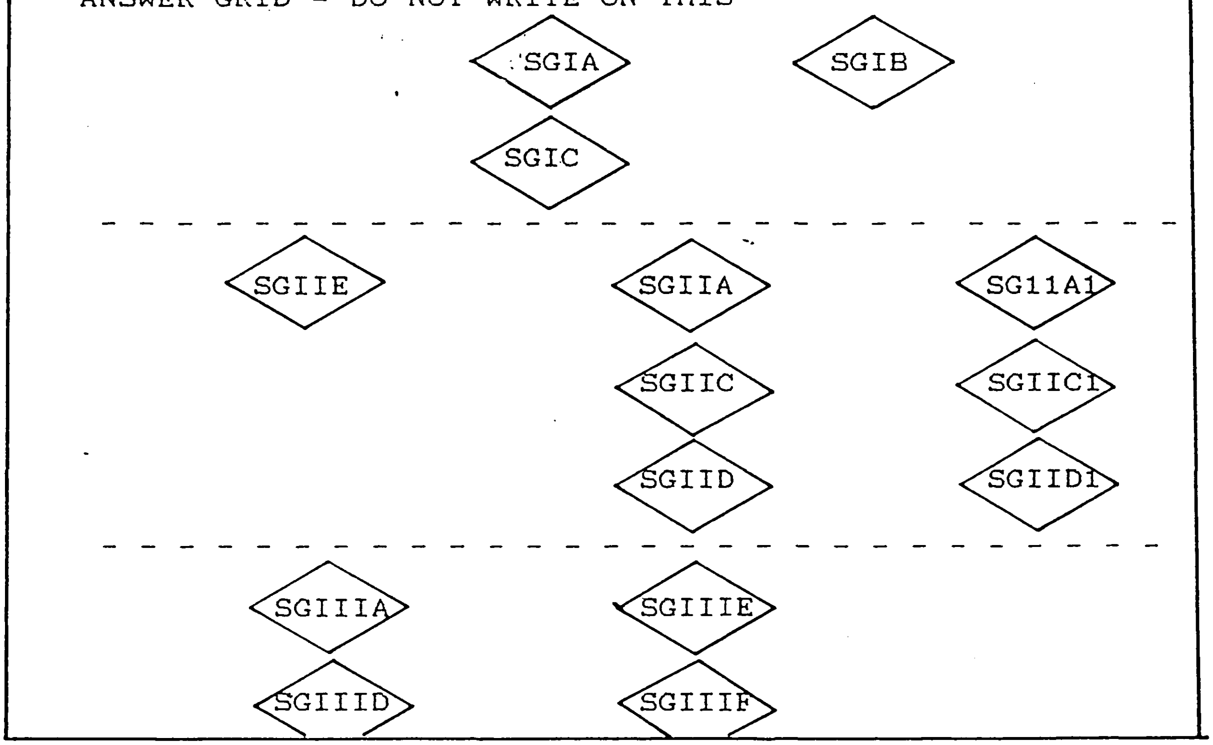
THERE IS NO TIME LIMIT BUT YOU SHOULD BE ABLE TO FINISH IN ONE HOUR

PERSONAL INFORMATION

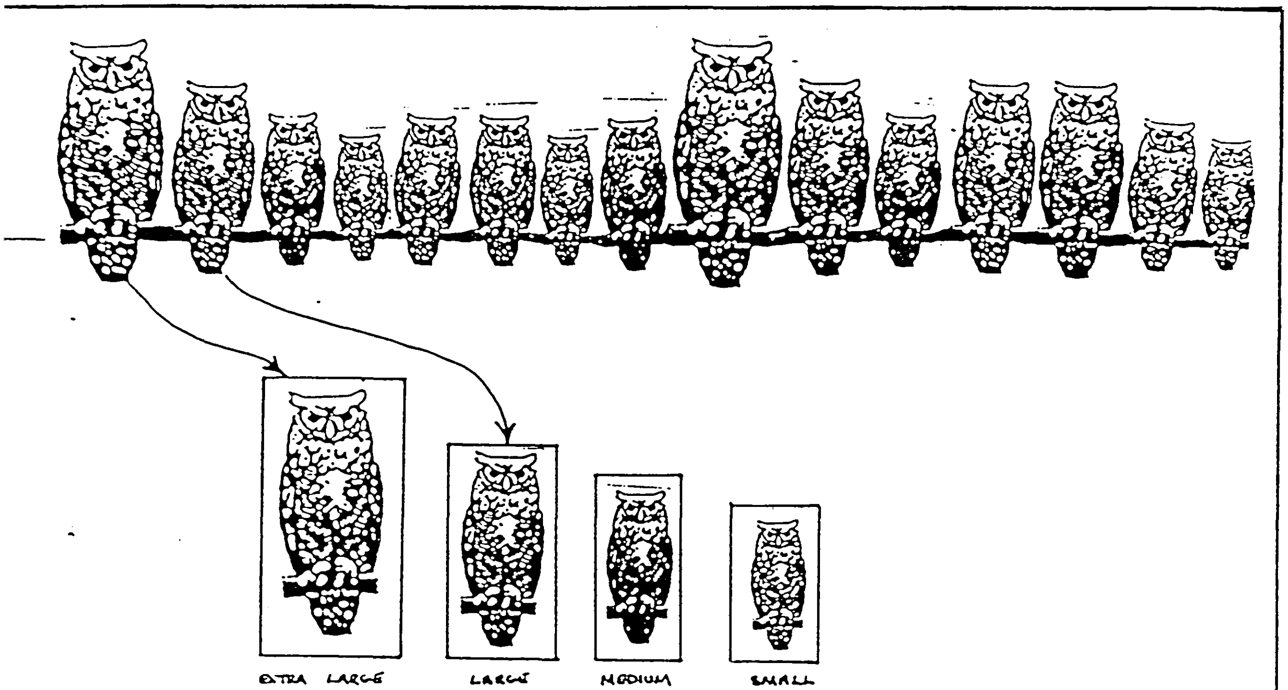
DATE OF BIRTH ....

AGE ON 1ST JAN 1991 ....

ANSWER GRID - DO NOT WRITE ON THIS



SG1B 1. In DATA A are pictures of owls. Draw arrows to show which size each owl is. The first two have been done for you.

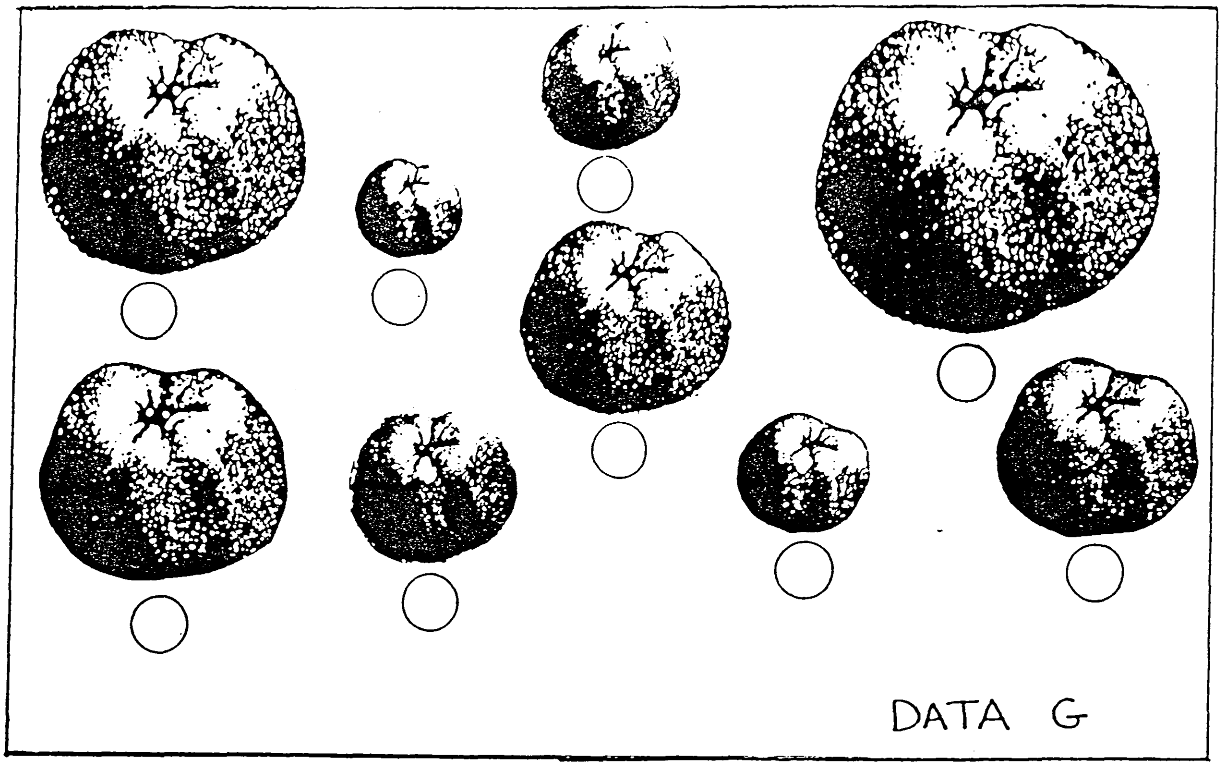


DATA A

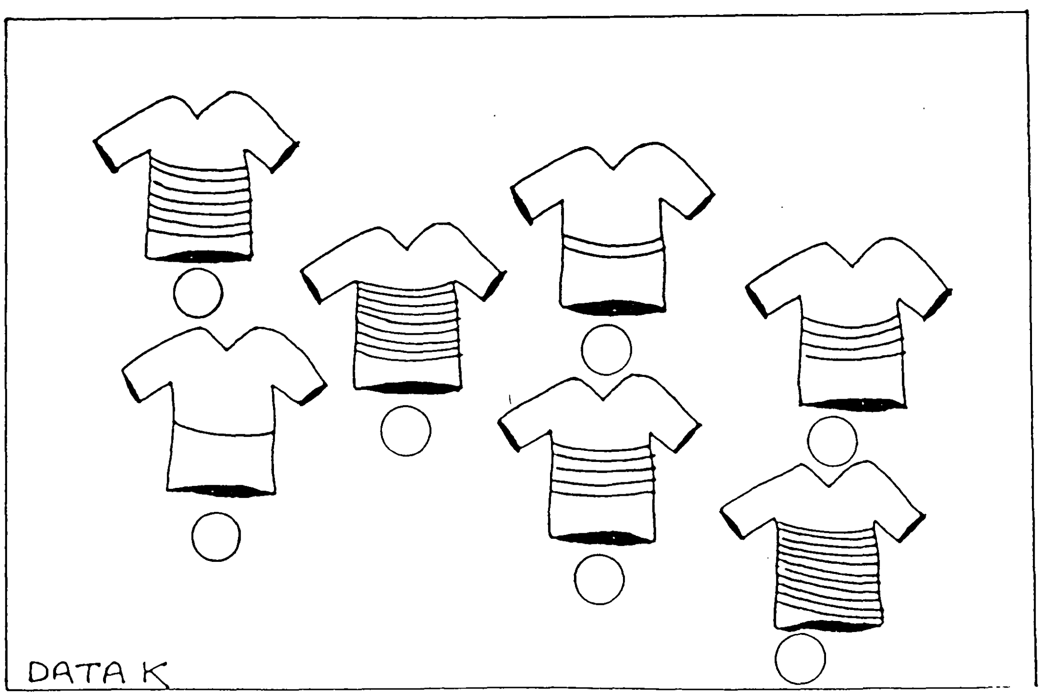
SGIIA 2. This table shows the number of owls of each size in DATA A. The first two have been done for you. Complete this table:-

OWLS	TALLY	TOTAL
EXTRA LARGE	I	
LARGE	I	
MEDIUM		
SMALL		

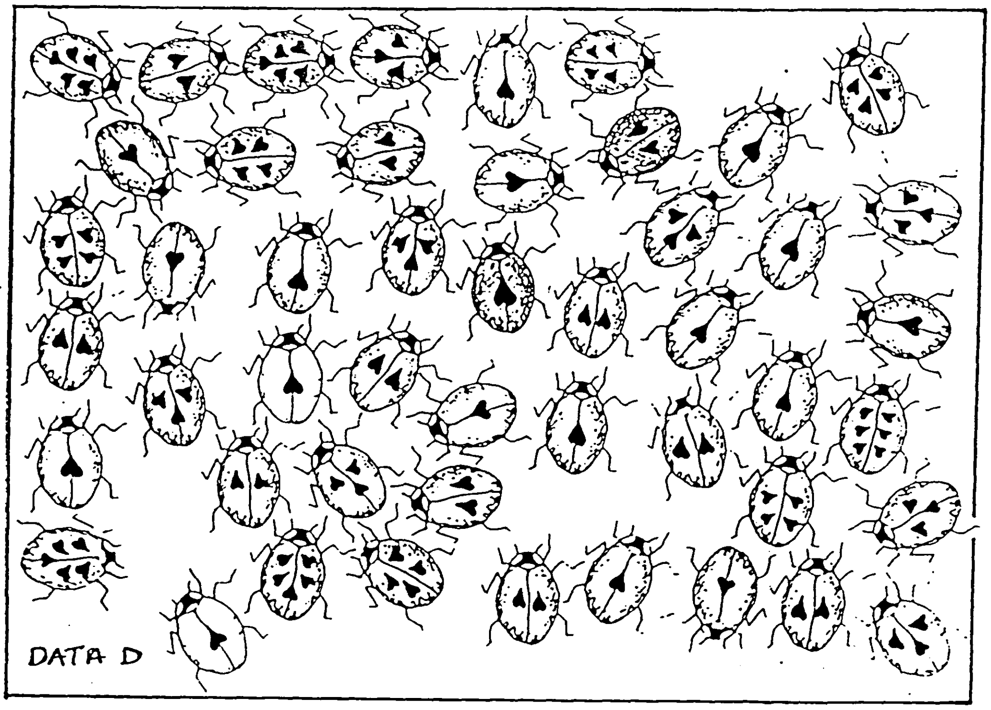
SGIC 3. In DATA G is a picture of 9 apples. Put a 1 under the biggest apple, 2 under the 2nd biggest apple and so on to number 9.



SGIA 4. DATA K shows some pictures of sweaters with stripes. Put 1st under the one with most stripes, 2nd under the one with next most stripes and so on.



SGIIA 5. DATA B shows a picture of some 'love-bugs' found by a scientist in the Amazon rainforest. Complete this table showing the number of spots:-



NUMBER OF SPOTS		TOTAL
1		
2		
3		
4		
5		
6		

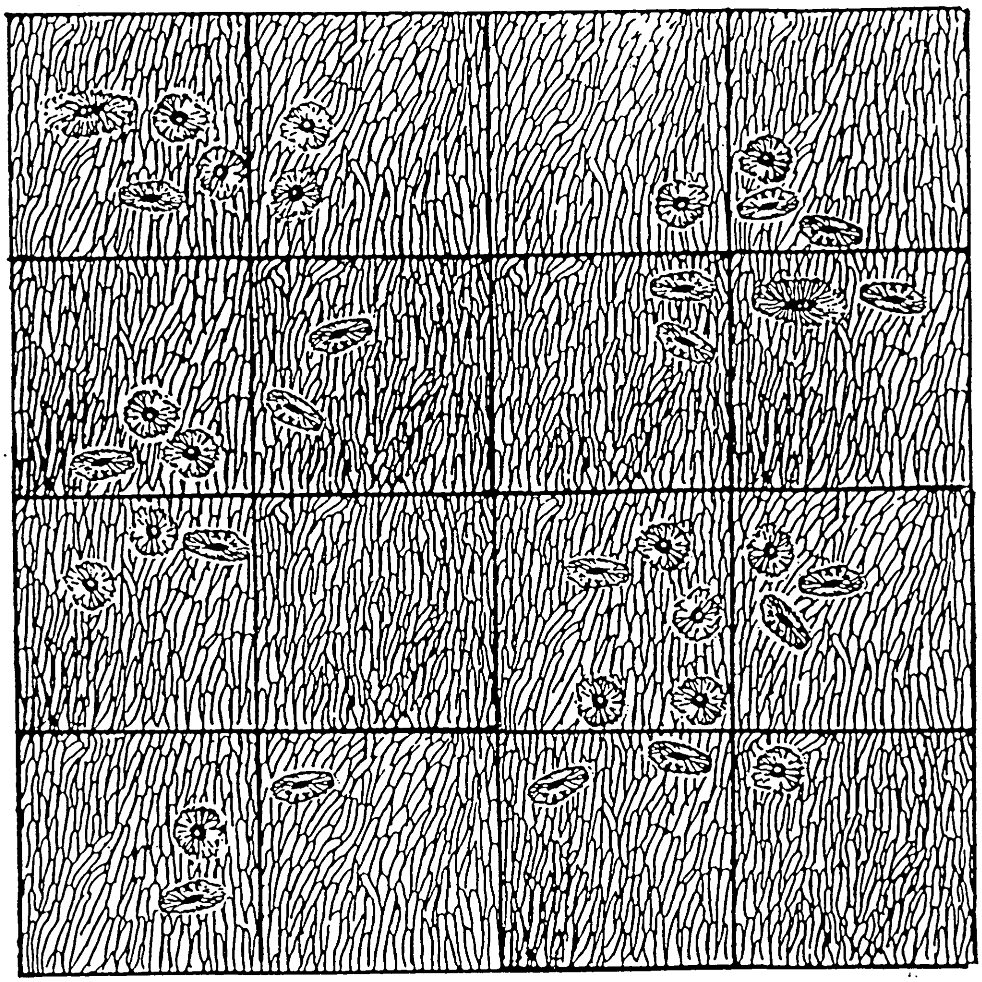
(a) How many love bugs have 3 spots ....

SGIIA1 (b) How many more three spotters are there than five spotters?................

SGIIA1 (c) Which are more common, two or three spotters?

SGIIB 6. DATA J shows part of a field with daisies growing on it. A botanist has marked out squares to count them. Use the table below to record the number of daisies in each square. The first square has been done for you.

DAISIES IN SQUARE	TALLY	TOTAL
0		
1		
2		
3		
4	I	
5		



DATA J

SGIIC 7. A class was given a test for which they were given a mark out of 50. The marks were:-
23, 34, 49, 16, 26, 23, 31, 41, 33, 27, 6, 19, 24, 34, 29, 15, 23, 18, 22, 38.

Fill in the table below to show their marks:-

MARKS	TALLY	TOTAL
1 - 10		
11 - 20		
21 - 30		
31 - 40		
41 - 50		

SGIIC1 (a) The teacher said she would be pleased if most people scored 21 or more. Would she be pleased with these marks? ....

SGIIC1 (b) In another class the scores were listed as:-

MARKS	TOTAL
1 - 10	6
11 - 20	10
21 - 30	7
31 - 40	5
41 - 50	2

Would the teacher have been pleased with this class on the same basis?
....

SGIIE (c) The classwere asked to make up their own tables. Several did it in different ways.

PUPIL A	PUPIL B	PUPIL C	PUPIL D
1 - 10	1 - 9	0 - 9	0 - 20
10 - 20	11 - 19	10 - 19	20 - 40
20 - 30	21 - 29	20 - 29	40 - 60
30 - 40	31 - 39	30 - 39	
40 - 50	41 - 49	40 - 50	

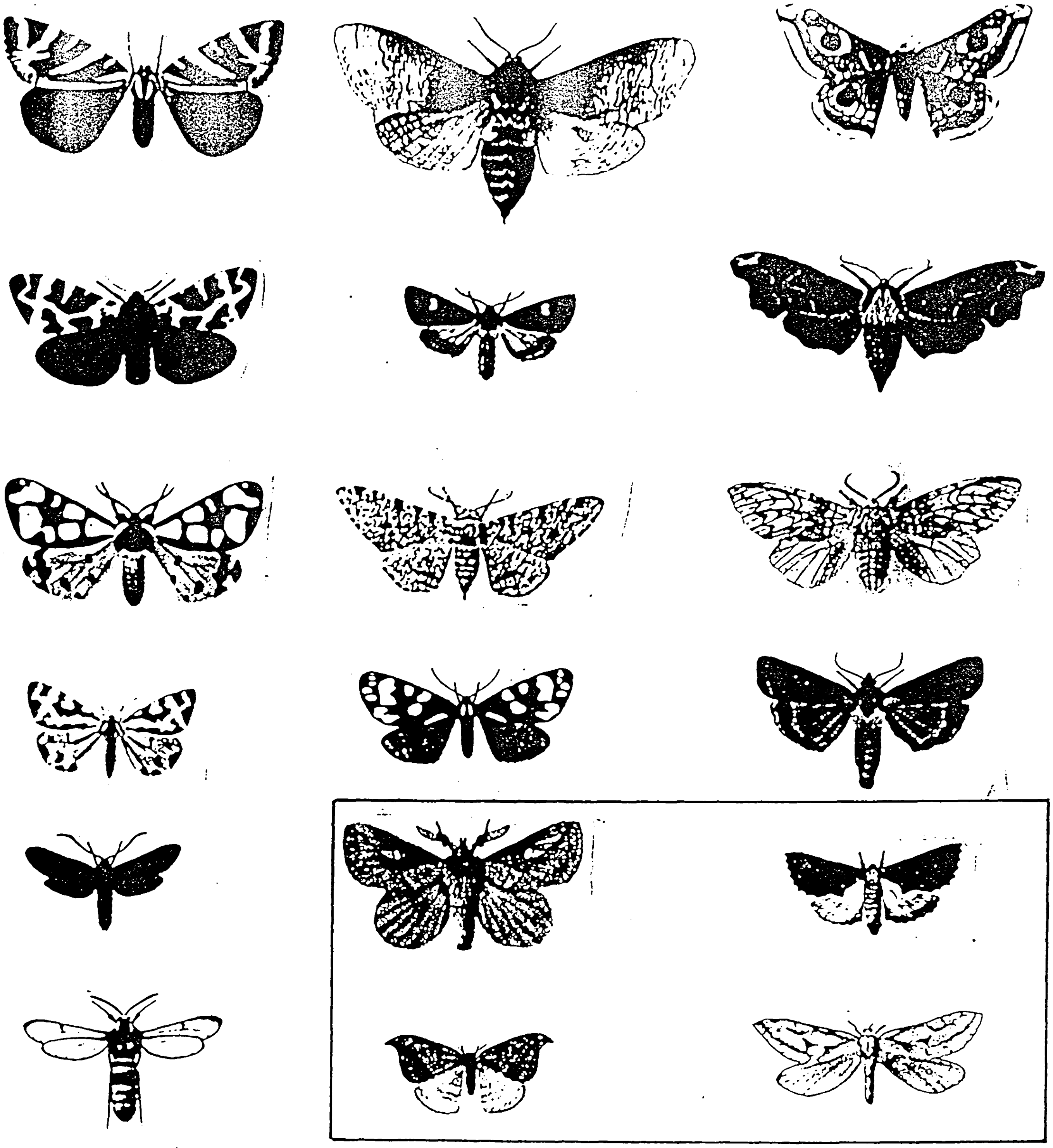
The teacher said that only one of the pupil's methods would work. Which pupil do you think it was? ....

SGIID 8. In DATA H are pictures of the most common British moths. Size is usually measured by 'wingspan' - the width between the tips of the wings at the widest place. Measure and record the wingspans of all the moths below. The four in the box have already been done.

WINGSPAN (cm)	TALLY	TOTAL	
2.0 - 2.9	I		
3.0 - 3.9	II		
4.0 - 4.9	I		
5.0 - 5.9			
6.0 - 6.9			

SGIID1 (a) How many moths have a wingspan between 3.0 and 5.9 cm? ....

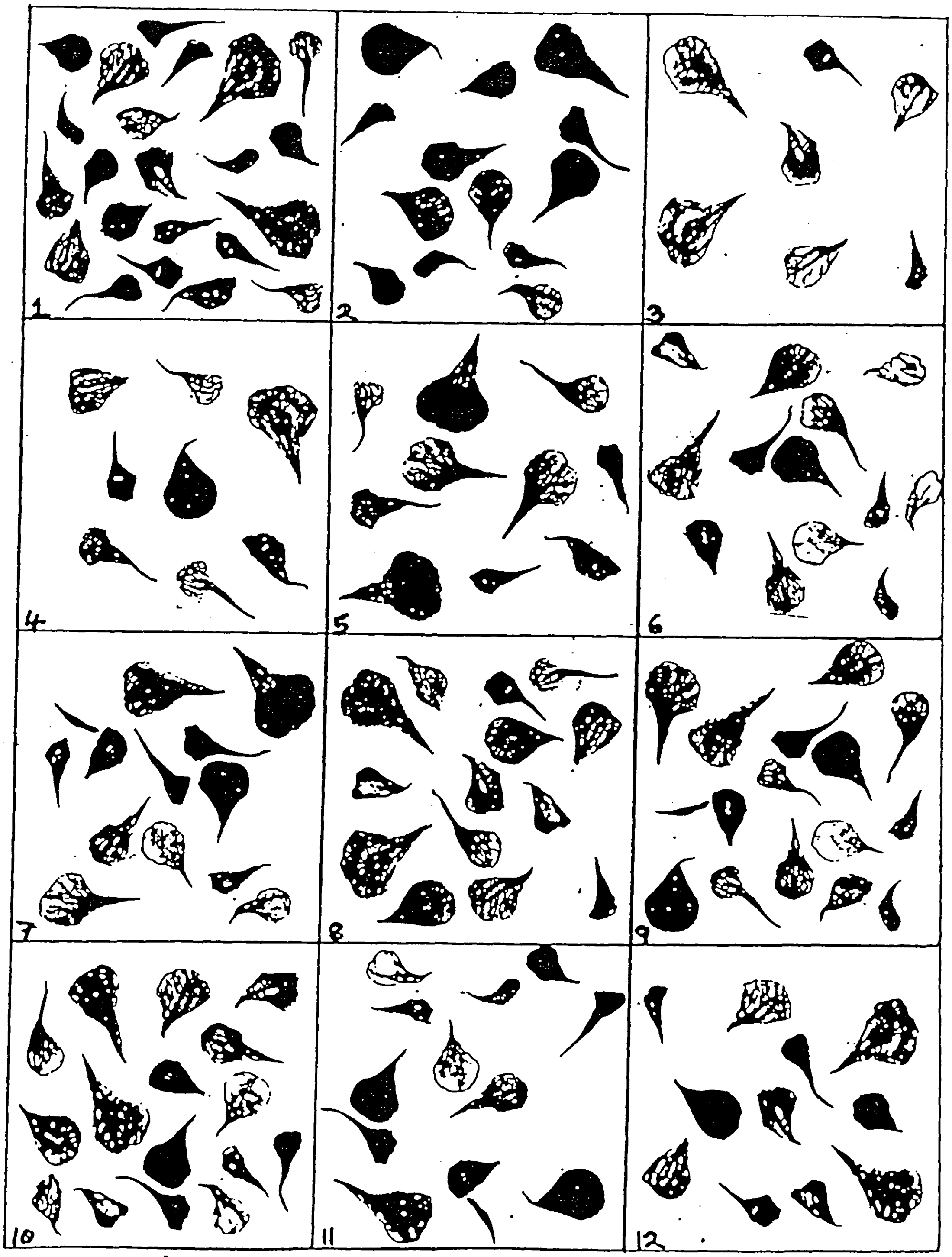
SGIIIF (b) Fill in the running total or cumulative total for each wing span group in the last column. That is the total in that group and all the groups below.



DATA H

SGIIB 9. A gardening expert was interested in the number of petals on geranium flowers. He took apart some flowers and laid them out as shown in DATA L. Complete this table:-

NUMBER OF PETALS		TOTAL
1 - 5		
6 - 10		
11 - 15		
16 - 20		
21 - 25		



DATA L

SGIID 10. A class have recorded their weights in kg. The weights are:-34.2, 44.6, 39.1, 52.3, 37.8, 46.7, 42.4, 43.8, 54.2, 58.7, 48.1, 44.4, 51.2, 47.0, 46.5, 60.8.

(a) Record this data in the table below:-

WEIGHT(kg)	TALLY	TOTAL	
30.0 - 34.9			
35.0 - 39.9			
40.0 - 44.9			
45.0 - 49.9			
50.0 - 54.9			
55.0 - 59.9			
60.0 - 64.9			

SGIIIF (b) The teacher wishes to work out a running or cumulative total for each weight range. Fill these in in the last column of the total.

SGIIIE (c) The teacher thought the the table was too spread out and decided to change the range of groups to cover 10kg. Use the table above to complete the new table below:-

WEIGHT kg	TOTAL
30.0 - 39.9	
40.0 - 49.9	
50.0 - 59.9	
60.0 - 69.9	

SGIIIE 16. DATA C shows some leaves from a laurel bush. Measure the lengths of the leaves from tip to tail and put the results in a table below using groups.

LENGTH OF LEAVES (cm)	TOTAL

DATA C

12. The same class also measured how tall they were very accurately in cm . They recorded the results as:-

HEIGHT(cm)	TOTAL
120.0 - 124.9	3
125.0 - 129.9	2
130.0 - 134.9	4
135.0 - 139.9	5
140.0 - 144.9	5
145.0 - 149.9	3
150.0 - 154.9	0
155.0 - 159.9	4

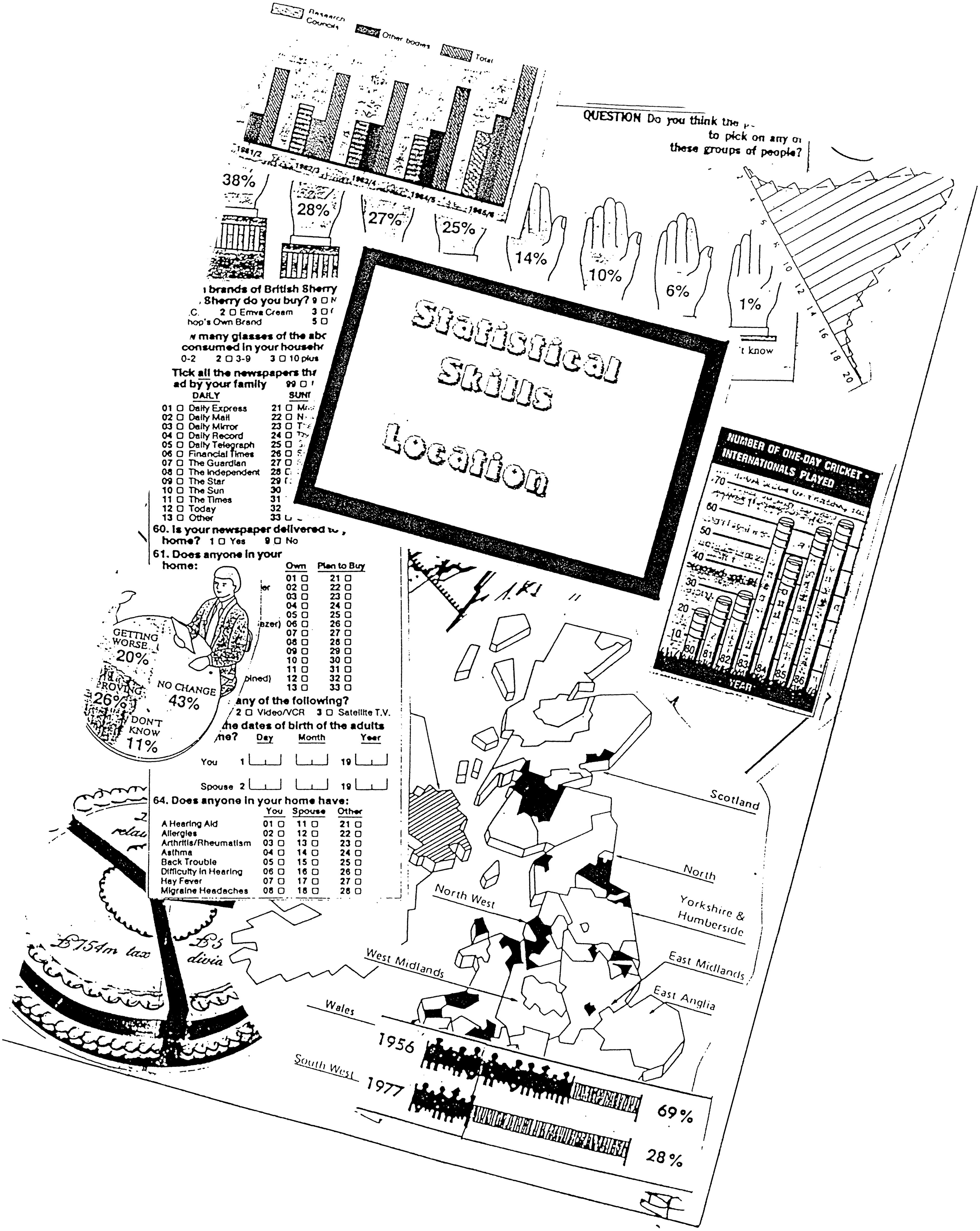
SGIID1 (a) Which were there more of, the tallest children or the smallest?
....

SGIIIA (b) Re-organise the table below so that each group contains a 10cm range:-

HEIGHTS(cm)	TOTAL

SGIIID (c) Which does there now appear to be most of, the tallest or smallest?............
Can you explain this?
....
....
....

END OF TEST



Statistical Skills
Location
QUESTION Do you think the
to pick on any of
these groups of people?
38%
28%
27%
25%
14%
10%
6%
1%
NUMBER OF ONE-DAY CRICKET INTERNATIONALS PLAYED
YEAR
GETTING WORSE 20%
NO CHANGE 43%
DON'T KNOW 11%
60. Is your newspaper delivered to home? 1 Yes 9 No
61. Does anyone in your home:
64. Does anyone in your home have:
A Hearing Aid
Allergies
Arthritis/Rheumatism
Asthma
Back Trouble
Difficulty in Hearing
Hay Fever
Migraine Headaches
Scotland
North
North West
Yorkshire & Humberside
East Midlands
West Midlands
East Anglia
Wales
South West
1956
1977
69%
28%

ANSWER AS MANY QUESTIONS AS YOU CAN - DO NOT WORRY IF THERE ARE SOME QUESTIONS YOU CANNOT ANSWER.

THIS TEST IS DESIGNED TO FIND OUT WHAT PEOPLE OF YOUR AGE KNOW

TAKE CARE WITH QUESTIONS YOU FIND EASY

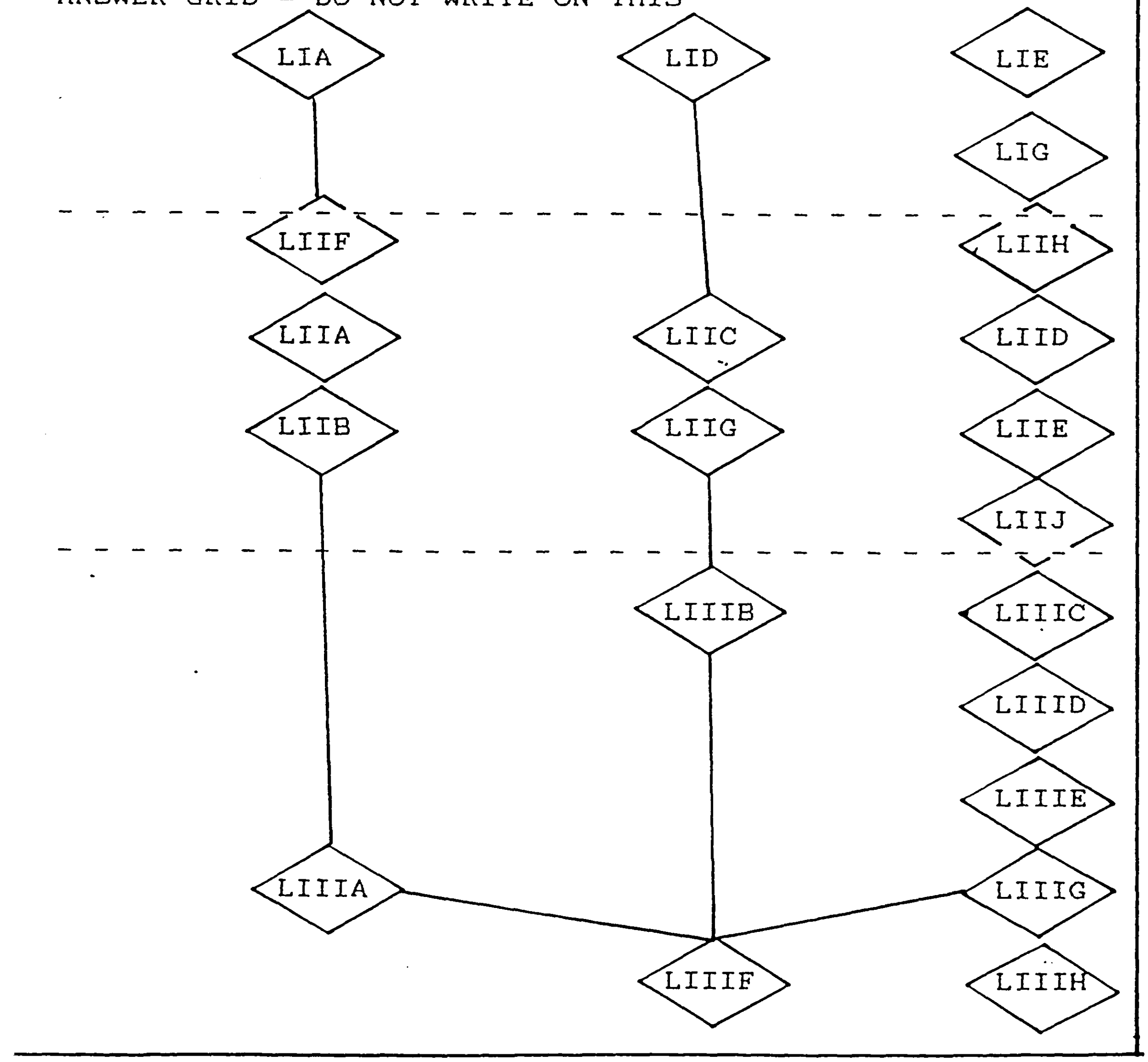
THERE IS NO TIME LIMIT BUT YOU SHOULD BE ABLE TO FINISH IN ONE HOUR

PERSONAL INFORMATION

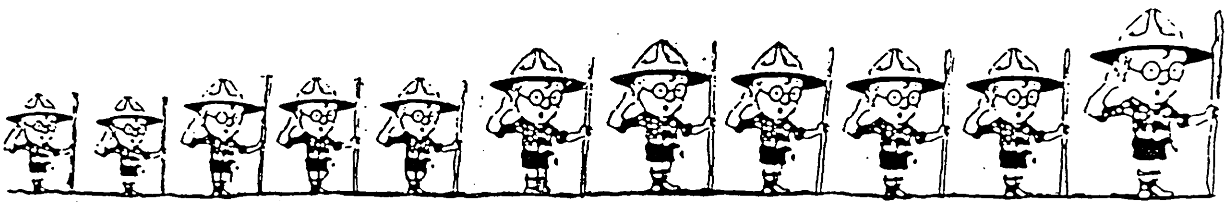
DATE OF BIRTH ....

AGE ON 1ST JAN 1991 ....

ANSWER GRID - DO NOT WRITE ON THIS

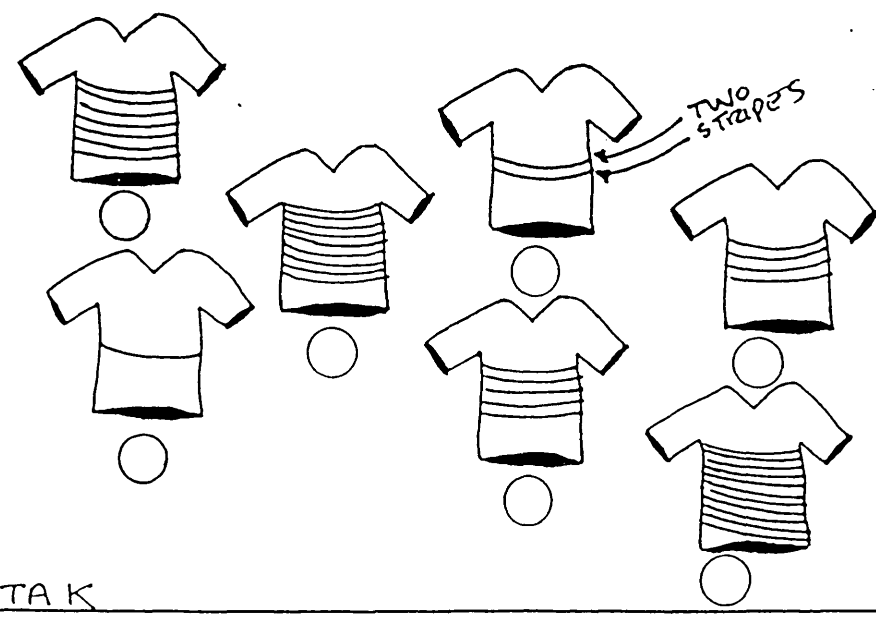


LIA 1. Some scouts were asked to stand in order of size and a picture of this is shown in Data A. Put a cross under the scout whose height is the middle size.



DATA A

LIIA 2. In Data K are some shirts with stripes. If they were put in order according to the number of stripes, how many stripes would the middle shirt have? ....

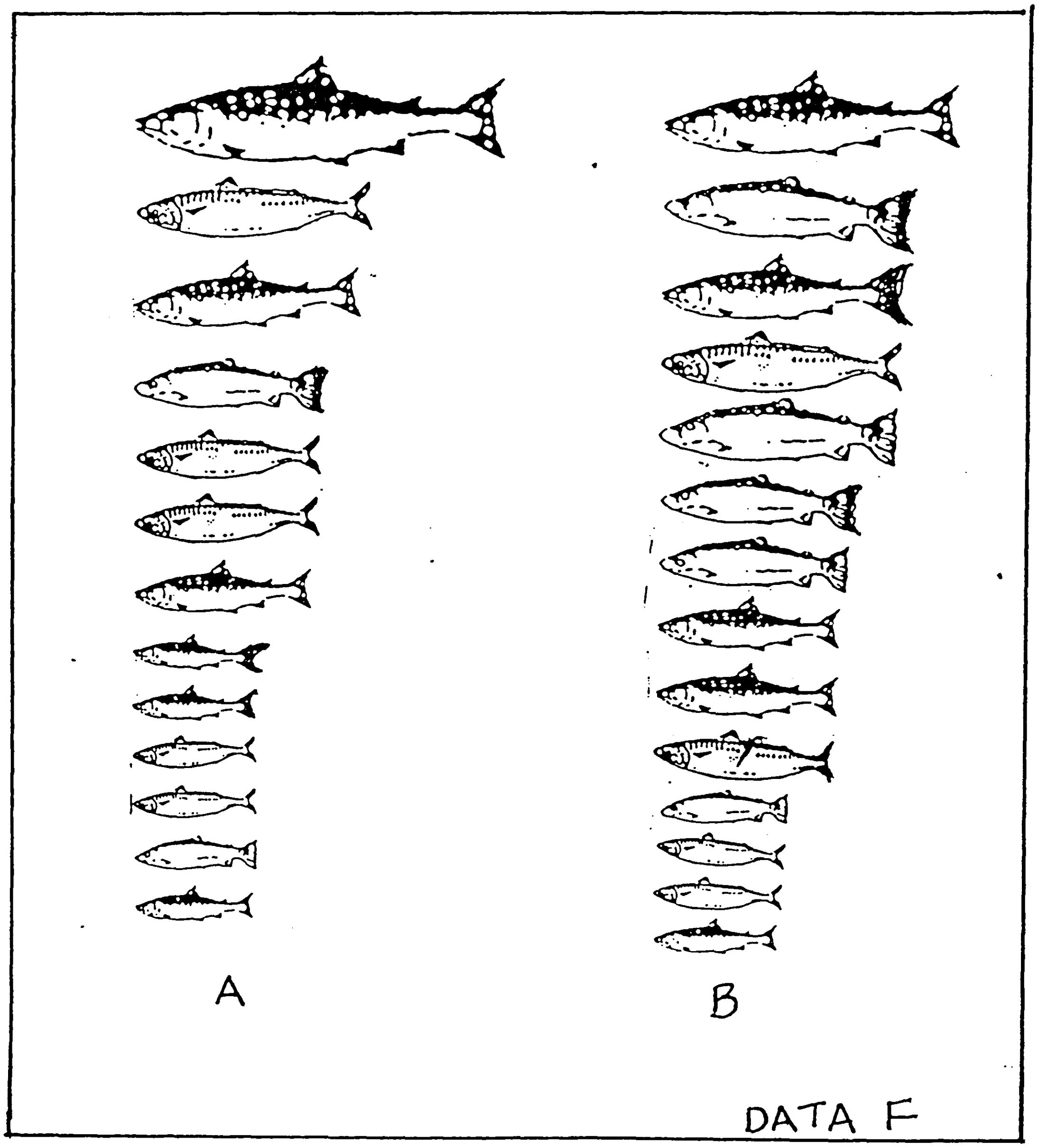


DATA K

3. Data F shows the fish caught by two boys one day.

LIA (a) Put a cross by the fish which is the middle one in size caught by boy A.

LIIF (b) Put a cross where you think the middle one for boy B is.

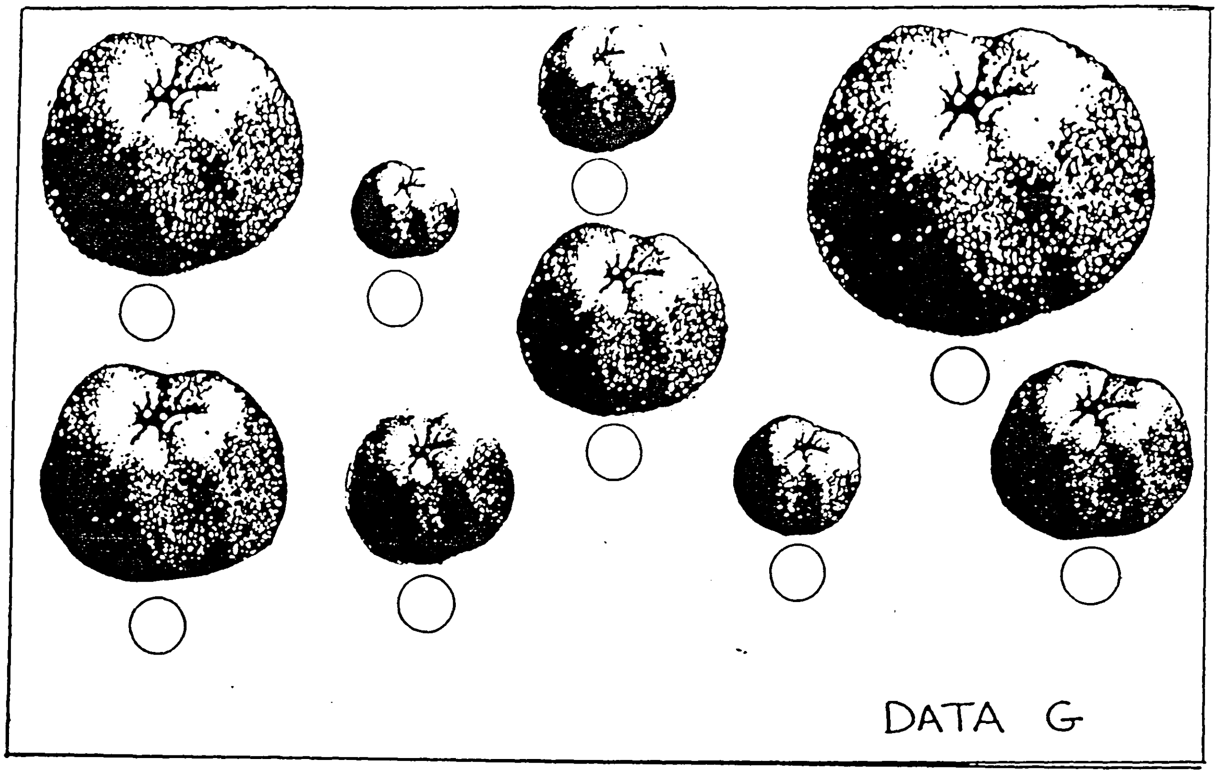


DATA F

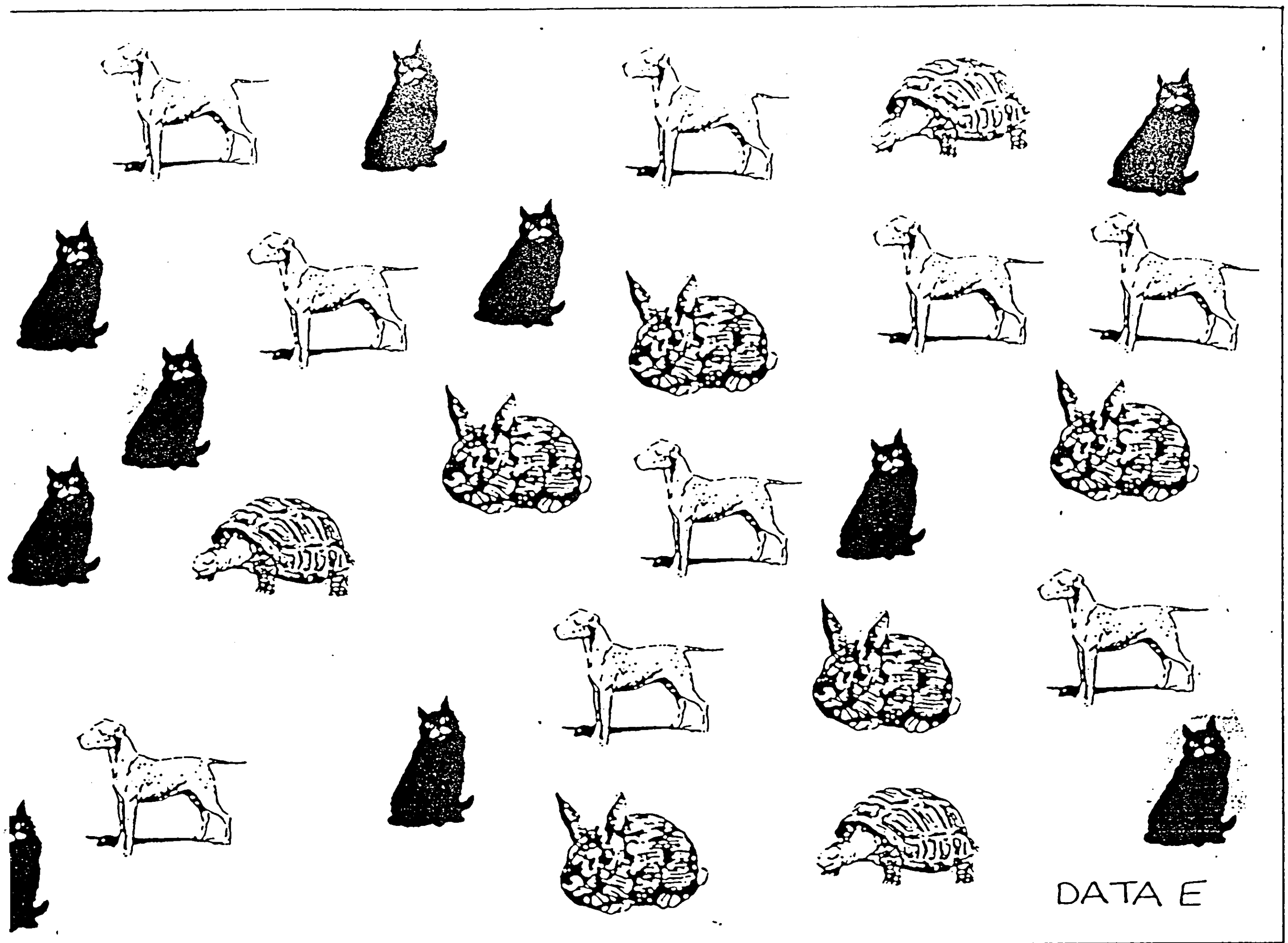
LIIA 4. Data G shows some apples. Write next to each apple 1 for the smallest, 2 for the next smallest and so on for all the apples.

(a) If they were laid out in order put an M in the circle underneath the middle one.

LIC (b) If there was another apple bigger than all the rest, which one would now be the middle one Put an X in the circle under this apple?



LID 5. Data E shows the pets owned by a class of children.
What is the most common pet in the class?
....



LIH 6. (a) Two classes did the same Mathematics test. Mrs Castle's group with 25 children got an average of 65. Mr Danes group had only got 15 children but had an average of 65 also. Which one of the following statements do you think is the fairest?

☐ The classes were just as good as the averages are the same

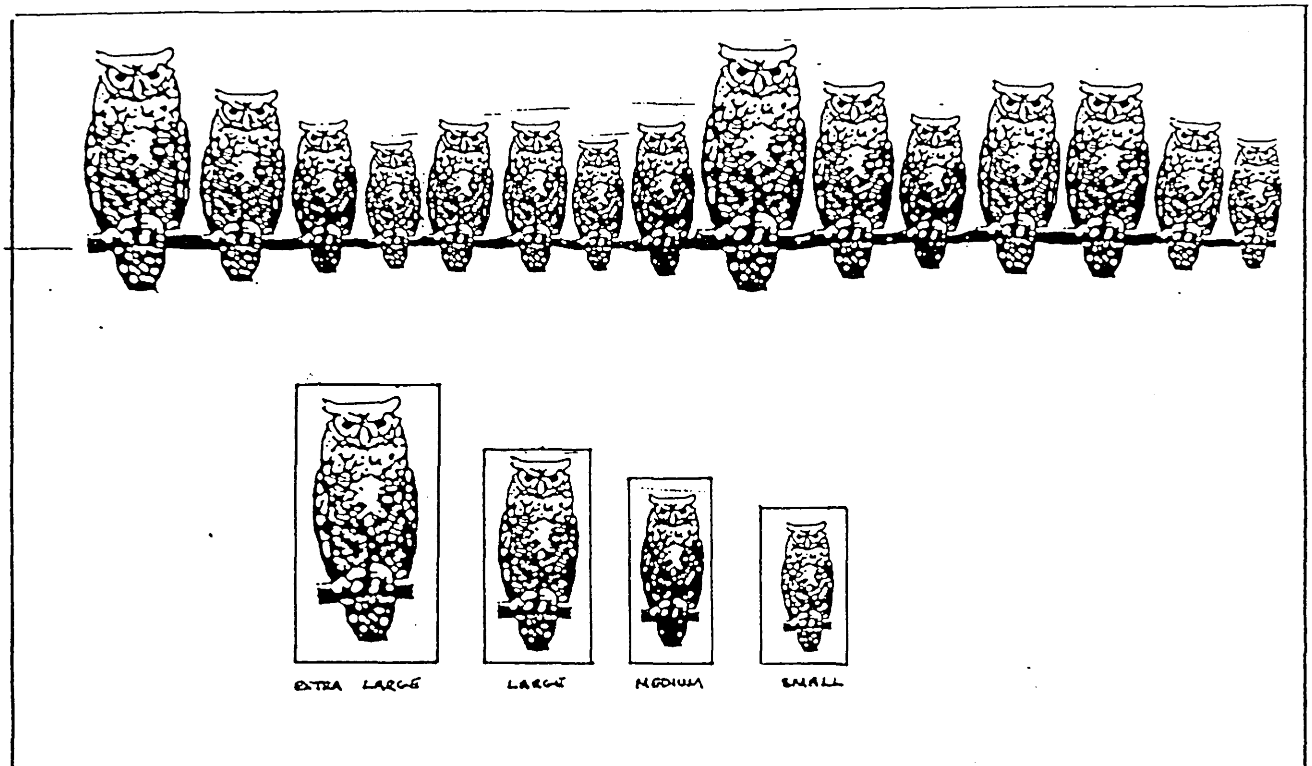
☐ Mrs Castle's group was better as the total score in the class must have been higher

☐ Mr Danes's group was better as they still managed to get the same average even though there were less of them.

TICK ONE BOX ONLY

LID 7. Data A shows some owls which are different sizes. What size is the most common?

....



DATA A

LIIG 8. (a) Two children were comparing the marks they had been given that term out of 10.

Jason had got: 8 9 8 7 7 8 7 Total 54

Sharon had got: 9 9 7 8 8 Total 41
(She missed two pieces of work when she was ill)

Which, if either, do you think got the better marks and why?

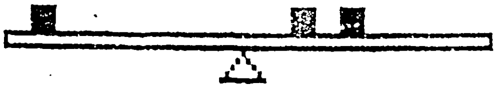
....
....
....

LIIE (b) Carl boasted that for his best five pieces of work he had an average(mean) of 9 marks. His friend knew though that he had been gven 0 for two other pieces of work. If the two 0's had been included would the average:-

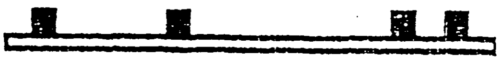
☐ stay the same ☐ be higher ☐ be lower

TICK ONE BOX

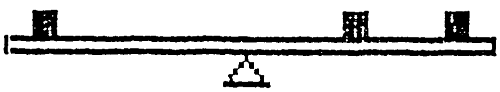
9. The following questions involve balances made of rulers. The black squares are blocks which weigh the same amount. In this picture the pivot, i.e. where we would balance the ruler, is shown by a triangle:-

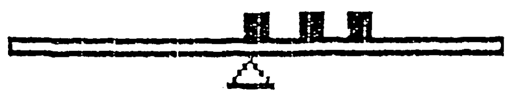


In these diagrams draw the pivot so the balance is level.



LIG (b) In these diagrams add one more weight to make the balance level.





10. Fill in the missing items of data in these examples:-

	DATA	MEAN(AVERAGE)
LIIIC	2, 5, 7, 9, ...	7
LIIIC	3, ..., 10, 14, 16	10
LIIID	2, 2, 5, 8, ...	6
LIIID	4, ..., 10, 12, 20	11

11. A group of schoolchildren were asked to write their shoe sizes on the blackboard. They wrote:-

4 7 8 5 6 7 5

LIID (a) What is the average or mean of the children's shoe sizes?
....
....
..

The whole class put their results in a table:-

SIZE	CHILDREN WITH THAT SIZE
4	3
5	5
6	7
7	8
8	6
9	2

LIIB (b) What is the middle or median size of shoes that the class takes?
....
....

LIIC (d) What is the mode or most common size? ......
....

LIIJ (e) What is the average(mean) size? Show how you got your answer...............................
....
....
....
....
....

12. Some children were asked how much pocket money they got. The amounts they gave were:-

2.50 2.50 5.00 3.00 2.50 3.00

LIID (a) What was the average or mean amount? ......
....
....
....

13. A group of children were asked how many children were in their family. The answers were;-

1, 2, 2, 1, 3, 5, 2, 4

LIIIE (a) What is the mean(average} number of children in families? ....

A class of children decided to collect the same data. They put their results in a table:-

CHILDREN	HOW MANY FAMILIES
1	4
2	8
3	7
4	1

LIIC (b) What is the most common number of children in the families in this class?

....

LIIJ (c) What is the average number of children for this group?...............................

....

....

....

LIIIE (d) One of the class said he had read in the paper that the average family in the UK had 2.2 children. In a discussion the following things were said:-

☐ This must be wrong as you cannot have 2.2 children

☐ Most families have exactly two children with a few having more

☐ Families have many different numbers of children but most are around 2.2

Tick the box next to the one you agree with most.

LIIIH 18.(a) In a youth club party the six girls bring on average 8 tins of drink. The three boys bring an average of 5 tins. How many tins on average did the nine members bring?

....

....

....

....

....

LIIIH (b) A boy has got an average mark of 9 for his five pieces of classwork. He has done three pieces of homework for which he has averaged only 5. What is his average mark for the eight pieces overall? ....

....

....

....

14. A class were asked to measure the lengths of some leaves they found to the nearest cm. They listed the results as follows:-

Length(cm)	Number of leaves
4	3
5	8
6	3
7	5
8	9
9	3

(a) What is the most common leaf size

....

LIIIB
LIIF
(b) The teacher asked the class whether it was fair to say that the leaves were typically 8cm long. These are some of the answers she was given:-

☐ Yes - because there were more 8cm leaves than any other size

☐ No - because there are far more leaves shorter than 8cm than longer

☐ No - because there are nearly as many 5cm leaves which we have ignored.

Tick the box that you agree with most.

15. A class measured their heights in centimetres and recorded them as follows:-

Height (cm)	Number
120.0 - 124.9	3
125.0 - 129.9	5
130.0 - 134.9	7
135.0 - 139.9	3
140.0 - 144.9	1
145.0 - 149.9	1

LIIIA (a) Find as accurately as you can the median(middle) height of the class.

....
....
....
....
....

LIIIG (b) Find as accurately as you can the average(mean) height of the class.

....
....
....
....
....

16. A class all weigh themselves in kilograms. They recorded the results as:-

Weight	Total
30.0 - 34.9	2
35.0 - 39.9	5
40.0 - 44.9	8
45.0 - 49.9	4
50.0 - 54.9	1

LIIIA (a) Find as accurately as you can the weight of the median(middle) pupil in the class.

....
....
....
....
....
....

LIIIG (b) Find as accurately as you can the average(mean) weight of the class.

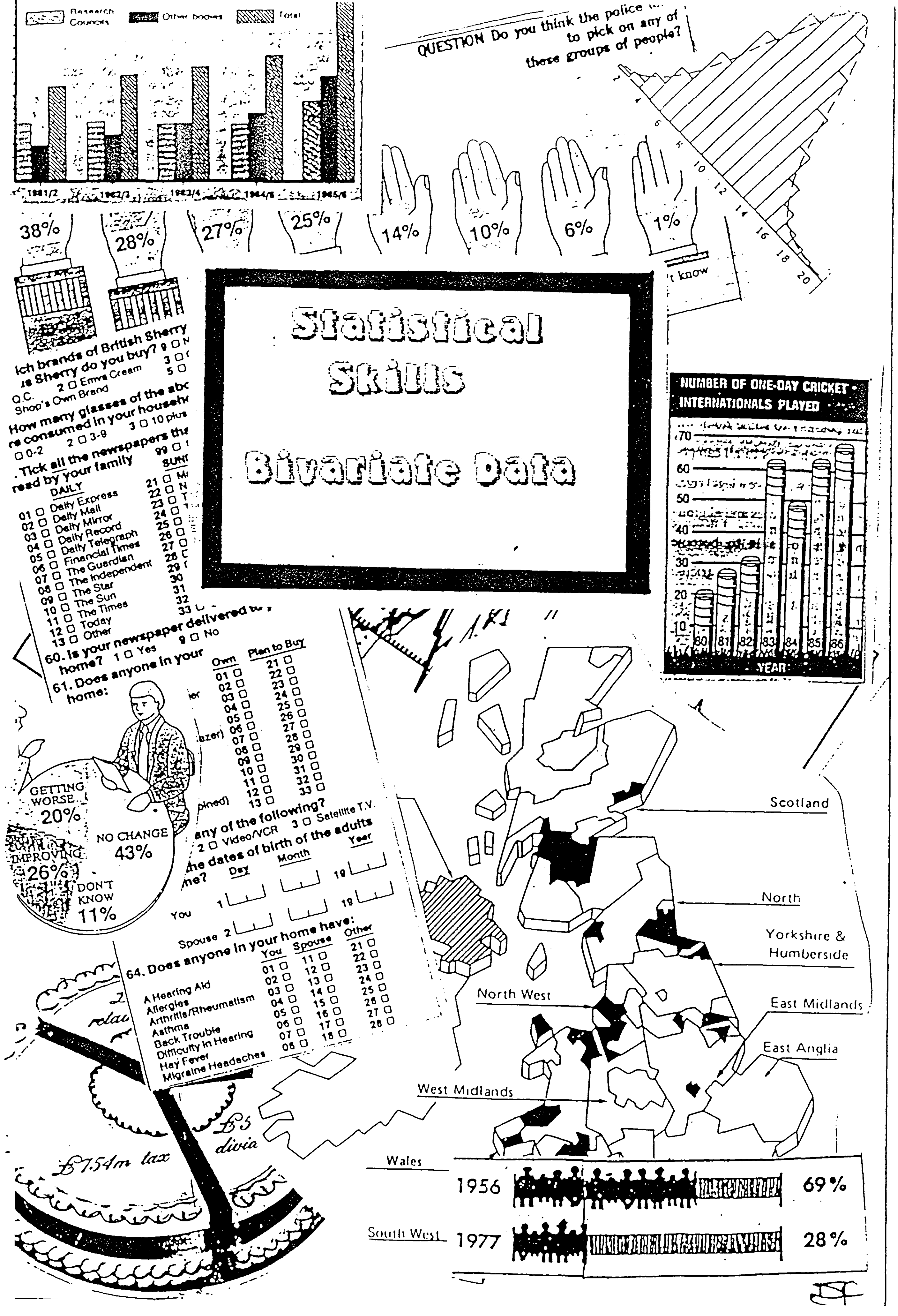
....
....
....
....
....
....

LIIIF 17. The information below was from a Government article showing how much people on YTS schemes earn. Do you think that the mean is a fair figure? If not what would you give as a fair figure and why?

....
....
....
....
....

Table 3 Income of YTS trainees (March 1984)

Income	Per cent of trainees:
£25.00 only	84
Over £25.00 up to £30.00	3
Over £30.00 up to £35.00	3
Over £35.00 up to £40.00	1
Over £40.00 up to £50.00	4
Over £50.00 up to £60.00	3
Over £60.00	2
	100
Mean £28.10	


Statistical Skills
Bivariate Data
Research Councils
Other bodies
Total
1981/2
1982/3
1983/4
1984/5
1985/6
QUESTION Do you think the police ... to pick on any of these groups of people?
38%
28%
27%
25%
14%
10%
6%
1%
NUMBER OF ONE-DAY CRICKET INTERNATIONALS PLAYED
YEAR
80
81
82
83
84
85
86
How many glasses of the ab
Tick all the newspapers th
read by your family
DAILY
01 Daily Express
02 Daily Mail
03 Daily Mirror
04 Daily Record
05 Daily Telegraph
06 Financial Times
07 The Guardian
08 The Independent
09 The Star
10 The Sun
11 The Times
12 Today
13 Other
60. Is your newspaper delivered to your home? 1 Yes 9 No
61. Does anyone in your home:
Own
Plan to Buy
64. Does anyone in your home have:
You
Spouse
Other
A Hearing Aid
Allergies
Arthritis/Rheumatism
Asthma
Back Trouble
Difficulty in Hearing
Hay Fever
Migraine Headaches
GETTING WORSE 20%
NO CHANGE 43%
IMPROVING 26%
DON'T KNOW 11%
Scotland
North
Yorkshire &
Humberside
North West
East Midlands
East Anglia
West Midlands
Wales
South West
1956
69%
1977
28%
£7.54m tax

ANSWER AS MANY QUESTIONS AS YOU CAN - DO NOT WORRY IF THERE ARE SOME QUESTIONS YOU CANNOT ANSWER.

THIS TEST IS DESIGNED TO FIND OUT WHAT PEOPLE OF YOUR AGE KNOW

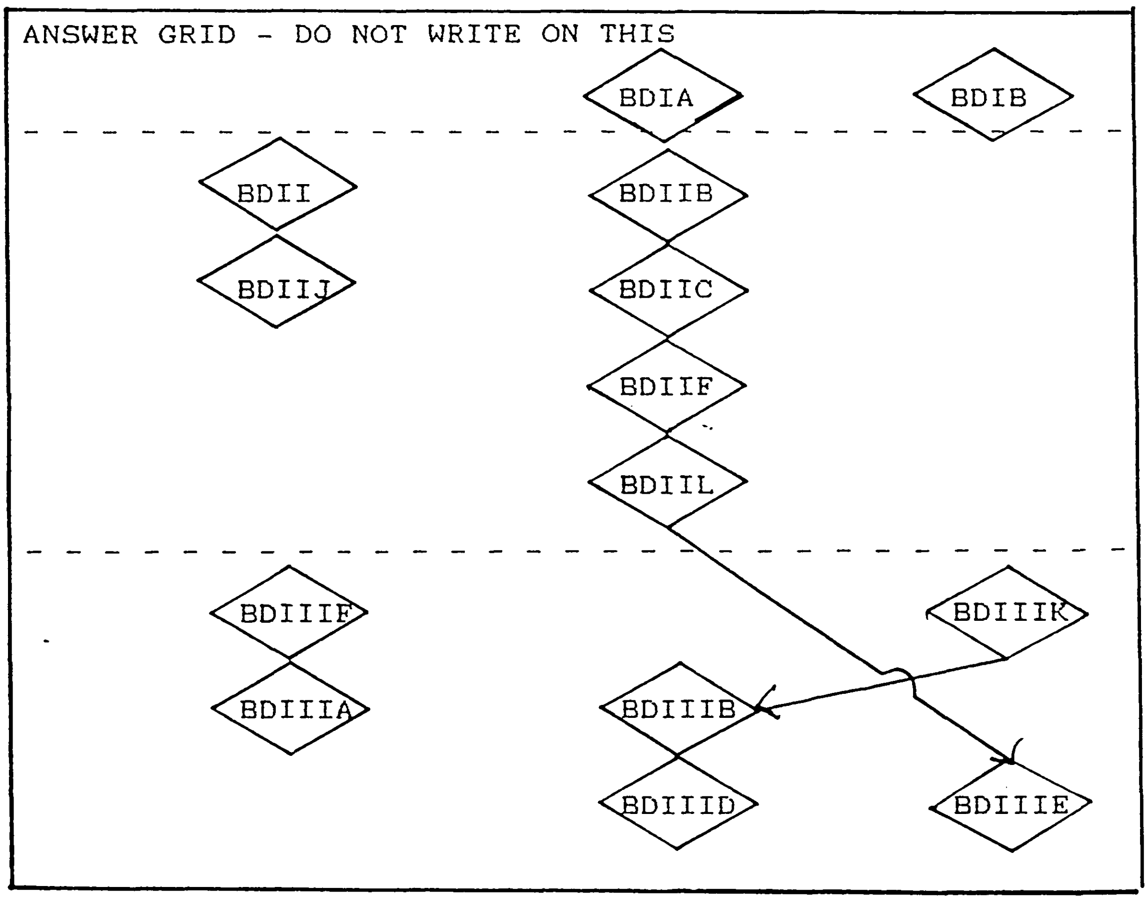
TAKE CARE WITH QUESTIONS YOU FIND EASY

THERE IS NO TIME LIMIT BUT YOU SHOULD BE ABLE TO FINISH IN ONE HOUR

<u>PERSONAL INFORMATION</u>

DATE OF BIRTH ....

AGE ON 1ST JAN 1991 ....



Bivariate Data

1. A newly wed couple want to rent a flat. They cannot decide whether to rent a one or two bedroom flat. They also want to look at furnished (furn.) and unfurnished flats. They cut some adverts out of a paper. These are shown in Data S.

BD1A (a) Do you think that one-bedroom flats are more likely to be furnished than two bedrooms? Or is it the other way round?..........................
....
..

BDIIA (b) Record the number of each type in the following table:-

	FURNISHED	UNFURNISHED
ONE BEDROOM		
TWO BEDROOMS		

14 CLASSIFIEDS

PROPERTY

A fully furnished 1 bedroomed self contained Flat, city centre, £60 p.w. Tel. Ashworth, Framingham Earl 5433.

A furnished Bedsitter to let on Mill Hill Road, Norwich, £125 p.c.m. Tel. Jones, Norwich 51932.

A large Bedsit, suit couple, Norwich centre. Tony, Norwich 764720.

A mod. 1 bed. Flat, Weavers Close, Stalham, £235 p.c.m. Simon, Hickling 304/mobile 0836 552075.

A superior 1-bed., furnished Flat, City Centre, Raleigh Court, ent. hall, fully fit. kit., lounge, bathroom, suit business couple or company let, £300 p.c.m., garage available. Hodgkinson. Tel. Norwich 621554, between 6-7 p.m.

A 1-bed. furnished Flat, £65 p.w. Richards, Norwich 701081.

day or evenings (0692) 650848.

HOLT. 2-bed. Flat, town centre, unfurnished, £265 p.c.m., excluding. Application forms from Bakers of Holt checkout. Tel. (0263) 712244.

HOLT. Period Cottage, furnished, 2

Bungalow, c/h, £400 p.c.m. Little, Sheringham 822201.

SHERINGHAM, 2 bed., unfurnished flat, £280 p.c.m. Tel. Bream, Sheringham 824798.

SHERINGHAM, 2 bed. unfurnished Second Floor Flat, from December 1st, g.c.h., parking space, £225 p.c.m. plus community charge, no pets. Tel. Howell, Sheringham

UNFURNISHED one-bedroomed Self-contained Flat at Lowestoft, £44 p.w., deposit required. Tel. Beccles 712777.

SUPERB large fully furn. 1 bed. Flat in City, ltd. kit., c.h., telephone, £310 p.c.m., incl. Tel. Stuart (0223) 68990 or (0480) 434313.

THORPE, furnished Room to let, use

A 1 bedrm. furnished Flat, near N & N, available immediately. Tel. M. French, Norwich 788821.

LUXURY 2-bed. Flat, unfurnished, 5 mins. city centre, couple preferred, £80 p.w. excl., deposit required. Tel. Carl, 0860-304054.

MOUSEHOLD centre, 1-bed. heated Flat, pkg., £45 p.w., excl. Billings Properties, Norwich 761576.

Home to let, near Stalham, suit single or professional couple, £50 per week. Mrs. Hoare. Tel. (069260) 777.

SHIPMANS TO LET

	PCM
NORWICH: West. 1-bed. Furnished Flat	£325
furnished	£350
HORNING: 4-bed. Bungalow, furnished	£400
NORWICH: 2-bed. Flat, unfurnished	£295

required. Eastfoolker Box No. N492H, Press Office, Norwich.

TWO-BEDROOM, unfurnished, Holt town centre Flat, for let. Application form from Iris Colman, Holt 712244.

UNFURN. clean warm 2 bed. terr. House, c.h., d.g., north city, £330

CROMER, elegant spacious 1-bed. 1st Floor Flat, central position, furnished, £250 p.c.m. Andrew, Cromer 511374.

HARLESTON, refurbished 2-bed. first floor Flat, £300 p.c.m., fully furnished. Tel. Harleston 852206.

Norwich 250001.

A nice, comfortably furn. 2 bed. c.h. Terraced House, Hall Rd., Norwich city end, w/machine, fridge, drinks bar, nice kit., £295 p.c.m. Tel. Reg, Norwich 250001.

A room in shared House off Thorpe Road, £35 per week inclusive, no D.S.S. Tel. Norwich 32220.

ACCOMMODATION available in large Farmhouse, on A144, 2 miles from Bungay, large single room and sep. lounge, fully furnished and c.h., share bathroom and kitchen, £45 p.w. inc. everything except meals. Tel. (098681) 579.

ATTLEBOROUGH sup. 2 bed. house, sat. TV, gge., gdns., all facils., £310 p.c.m.

AYLSHAM town centre, furnished Flat, 2 beds., Economy 7 heating. Tel. Rowlands, Aylsham 733273 (days), 732534 after 6 p.m.

420.

RALEIGH Court. Superior City Centre unfurn. 1-bed. Flat with pking. space, £295 p.c.m. References required. Tel. Brooks, Norwich 504353.

FAKENHAM. One-bedroom Flat to rent, (town centre), furnished, gas c/h., d/g. Smith (0328) 862856.

SELF-CONTAINED 2 bed. furnished Flat in Wymondham, £65 p.w. Tel. Quarrie, (0223) 843227.

DATA S

3. A biologist is looking at the growth of tadpoles. These get larger with time. She takes a photograph of twenty of them. This is shown in DATA R. She writes underneath them their age in weeks and days. (2:1 is two weeks one day old). She starts to list the age (in days) and the length (in cm) of the tadpoles. She does those above the line as follows:-

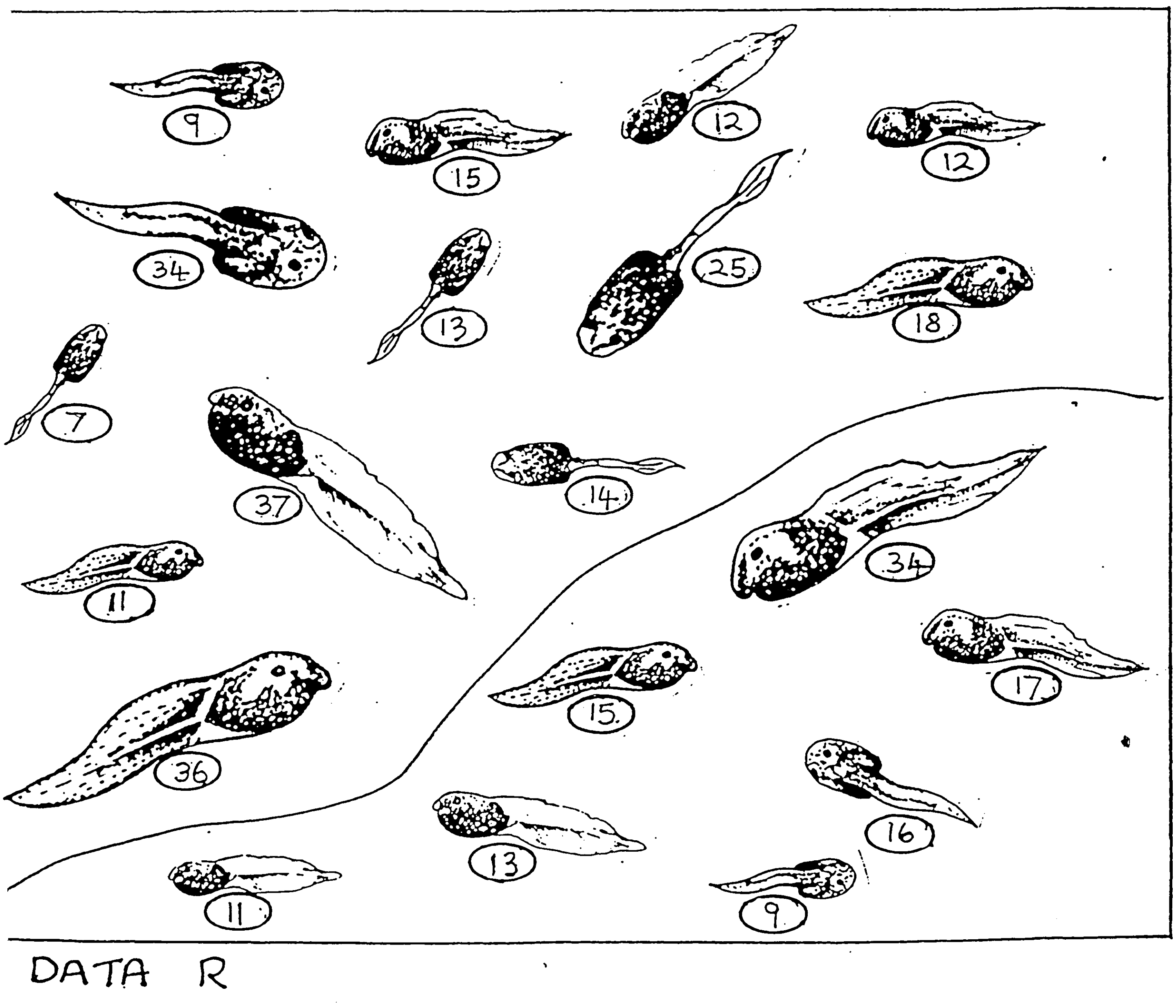
AGE LENGTH

(9, 2.2) (15, 3.4) (12, 3.3) (12, 2.8)
(34, 4.4) (13, 2.8) (25, 4.6) (18, 3.6)
(7, 2.4) (37, 5.3) (14, 3.1) (36, 5.5)
(11, 2.8)

DIIB (a) Write down the remaining seven pieces of data in the same way:-

(,) (,) (,) (,) (,)

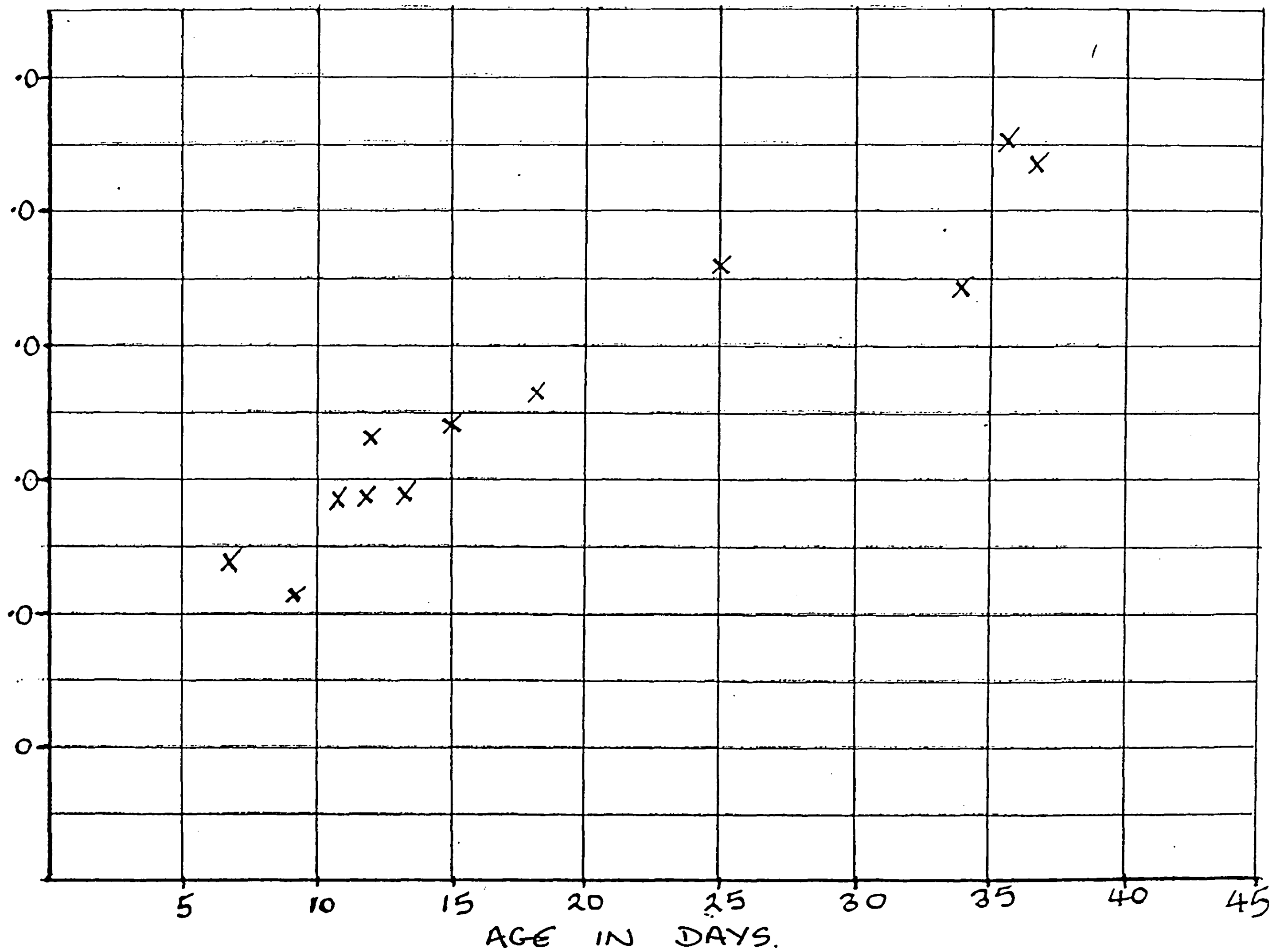
(,) (,)



DATA R

graph using one point for each tadpole. The ones listed above have already been done on the graph below.

BDIIC Add your seven values to the graph.



BDIIF (c) She thinks that it would be a good idea to draw a line to show the connection between age and length on the graph. Draw a line on the graph which you think best shows this.

BDIIIE (e) What length would you expect a tadpole 6 weeks old to be?

....

BDIIIF (f) Someone suggests that it would be neater to put the data in a table. She decides to do this for the first 13 pieces of data as follows:-

Length (cm)	1:0-1:6	2:0-2:6	3:0-3:6	4:0-4:6	5:0-5:6
2.0-2.9	~~IIII~~				
3.0-3.9	I	III			
4.0-4.9			I	I	
5.0-5.9					II

Age (Weeks:days)

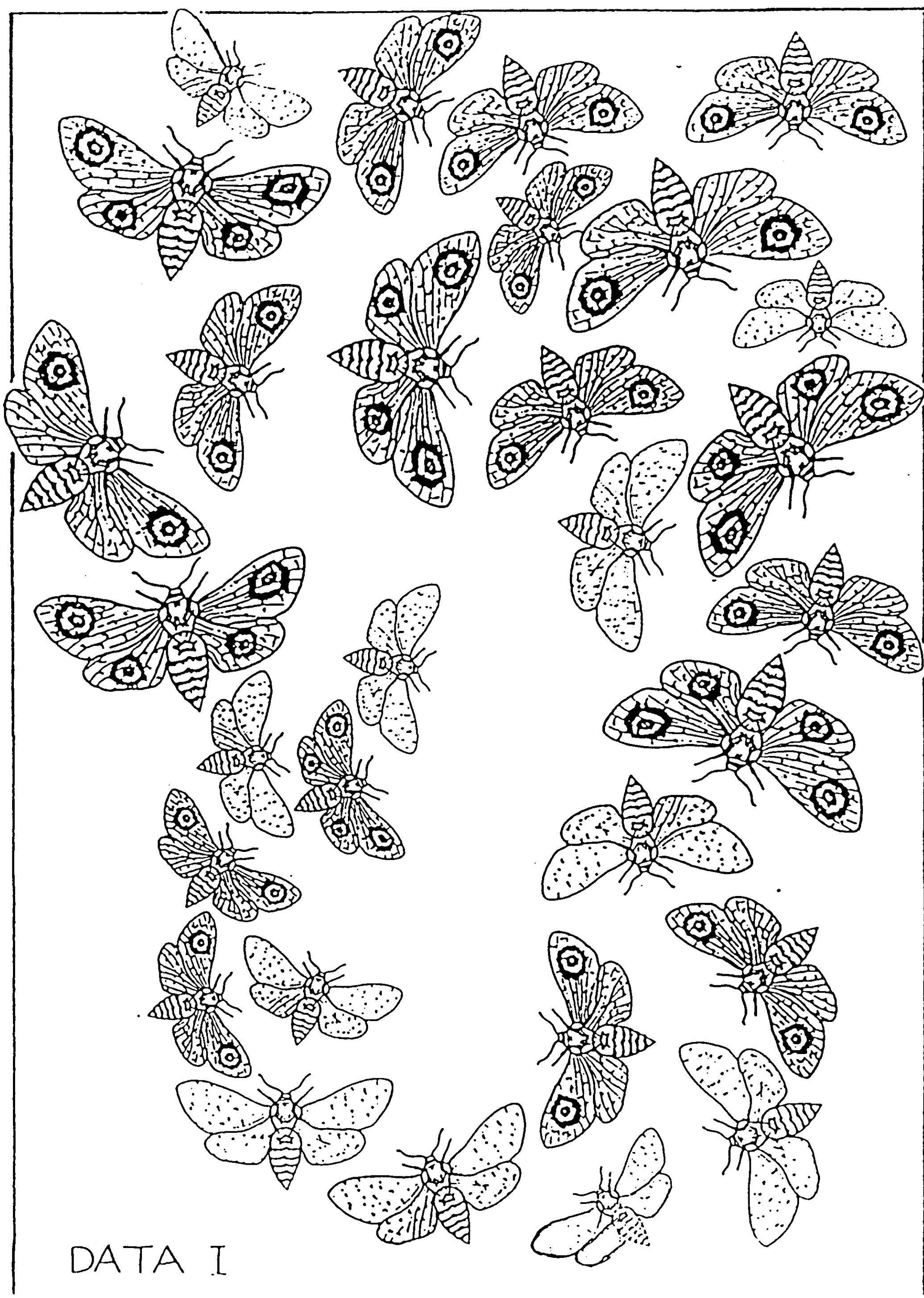
Add your seven results to the table.

3. A scientist is examining a new breed of moths. Data I is a picture of some moths of this breed. There are three sizes, small, medium and large. They have either four, two or no spots.

BDIA (a) Do you think that bigger moths have more spots? Or is it the other way round?..................
....

BDIIA (b) Write down how many of each type there are in the table below:-

	SMALL	MEDIUM	LARGE
NO SPOTS			
TWO SPOTS			
FOUR SPOTS			



BDIIJ 4. A few years ago, before everyone had to wear seat belts, this table was printed in a newspaper.

Which of the following statements do youthink most clearly sums up what thistableshows:-

☐ There is virtually no difference in how likely you are to be killed whether you are wearing a seatbelt or not

☐ You are more likely to be killed in an accident if you were wearing a seatbelt

☐ You are more likely to be killed if you were not wearing a seatbelt.

TICK ONE BOX ONLY

DRIVERS AND FRONT SEAT PASSENGERS
IN CARS AND LIGHT VANS

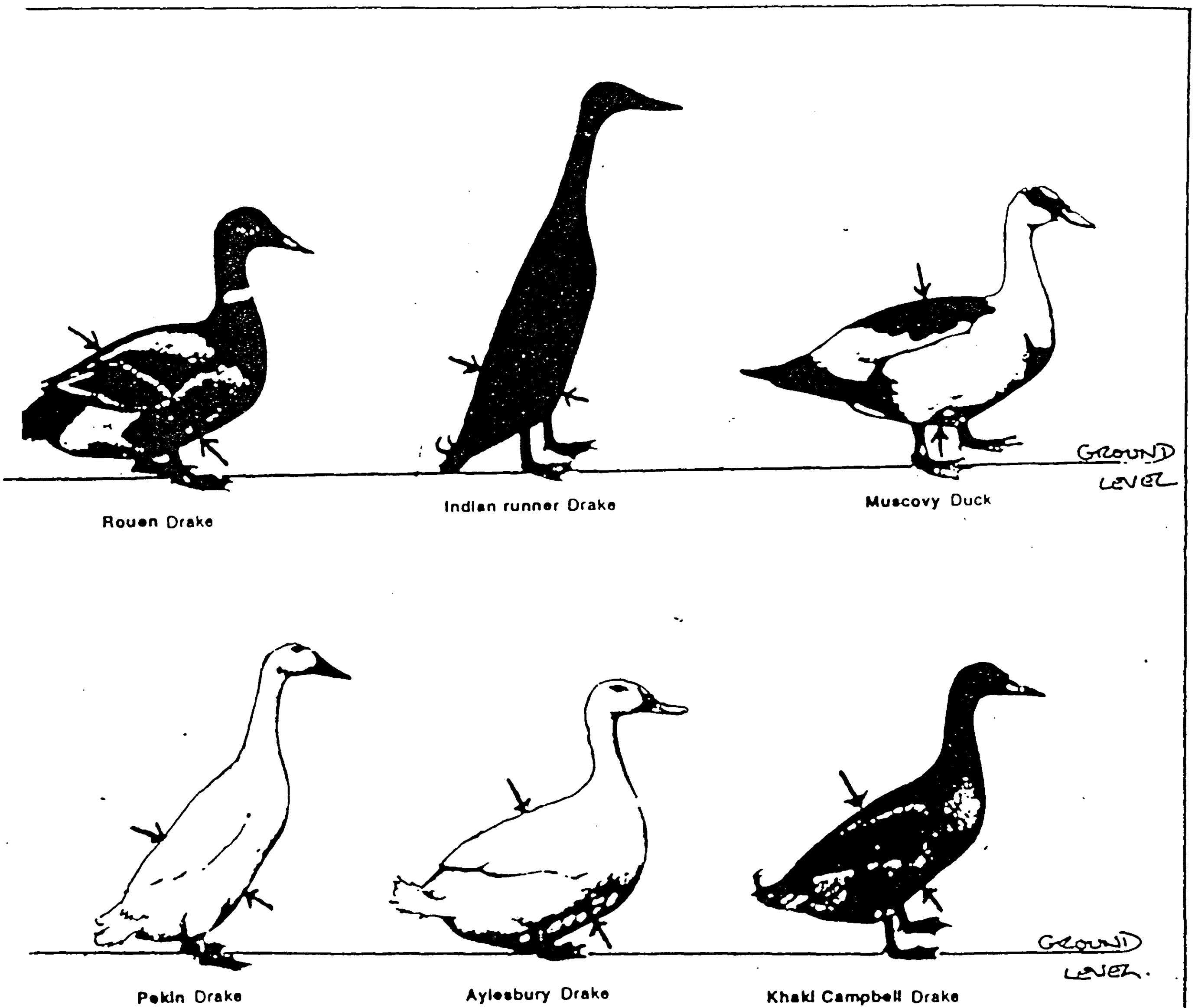
	CASUALTIES			
Safety Belt	Killed	Serious	Slight	Total
FITTED AND WORN	4	72	447	523
FITTED BUT NOT WORN	62	612	2543	3217

5. DATA T shows the six most common native British ducks.

BDIIB (a) Measure their heights in cm and their 'waists' in cm as shown by the arrows. Record the information in the brackets below (the first one has been done):-

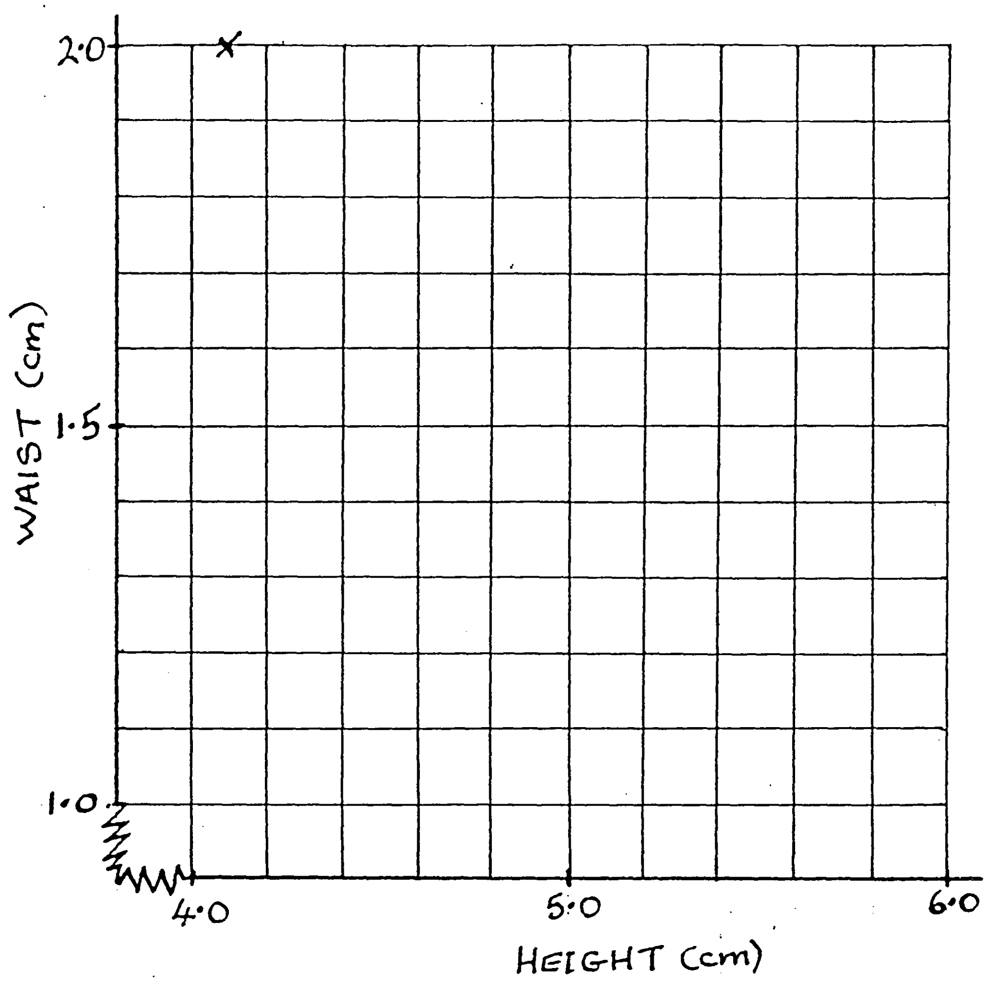
(4·1, 2·0) (,) (,) (,)

(,) (,)



DATA T

The first one has been done for you.



BDIIF (c) Draw a line on the graph which you think best shows any link between height and waist.

BDIIL How well do you think your line shows the link between height and waist?

....

....

....

....

BDIIIE (d) What waist measurement do you think a duck 8cm tall would have?.............

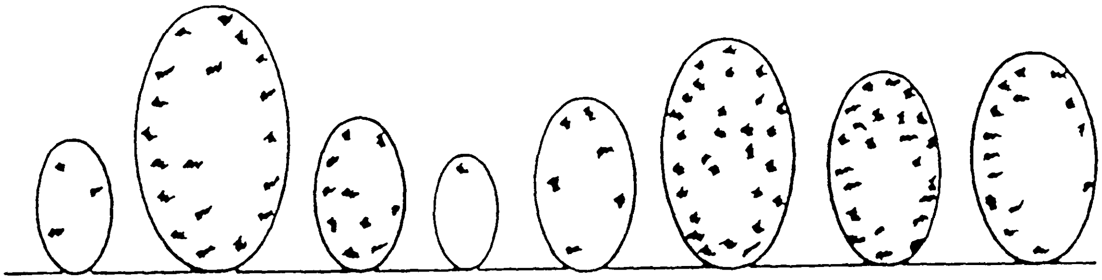
BDIIID (e) Tick the box of the sentence below which you think best describes what the data shows:-

☐ Taller ducks tend to have bigger waists.

☐ Taller ducks tend to have smaller waists.

☐ Taller ducks are just as likely to have small or big waists.

BDIB 6. A farmer has eight eggs from his chickens as shown in DATA U. He wants to grade them according to size and how speckled they are.
(a) In the size row put 1st under the biggest and so on. Do the same for the one with the most specks to the least.



SIZE								
SPECKS								

DATA U

(b) Tick in the box next to the statement which you think is shown in these results:-

☐ Bigger eggs are more speckled on the whole.

☐ Smaller eggs are more speckled on the whole.

☐ The number of specks is in no way connected with size.

7. A doctor is trying to see if a new drug cures a disease. He has 48 patients with the disease and he gives 18 of them the drug. The rest are given nothing. He then sees how many people in each group are better after one week. He lists the results as:-

	BETTER	NO BETTER
GIVEN DRUG	10	8
NOT GIVEN DRUG	13	17

BDIIJ Which of the following statements best describes what the results show:-

☐ Patients who were not given the drug were more likely to get better

☐ Patients were just as likely to get better whether they were given the drug or not

☐ Patients were more likely to get better if they were given the drug

8. A biologist is measuring the lengths and weights of a small rodent group. He lists the results as follows:-

Length(cm)	14.3	14.8	16.1	15.5	15.7	14.9	15.3	15.2
Weight(gm)	34	38	48	41	43	40	42	40

BDIIIF (a) Record the data in the table below:-

Length (cm)	30-34	35-39	40-44	45-49
14.0-14.9				
15.0-15.9				
16.0-16.9				

Weight(gm)

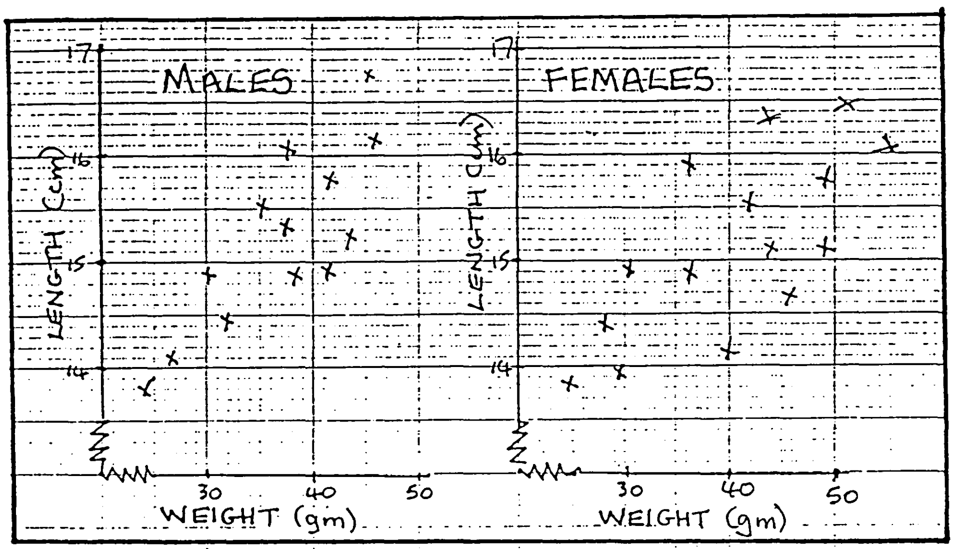
BDIIIA (b) What does this table show about the connection between length and weight?

☐ A: Longer rodents tend to weigh more

☐ B: Shorter rodentstend to weigh more

☐ C: Long rodents are just as likely to be heavy as short rodents

The biologist collects some more data. This time he keeps the information for males and females seperate. He draws a graph of the results for each. These are shown here.



BDIIL (c) Which of the statements A,B or C above applies to female rodents? .....

BDIIIB (d) Which of the following statements best describes thow strong the links between length and weight are:-

☐ The link is stronger in males than females

☐ The link is about the same for males as females

☐ The link is weaker in males than females

9. An insurance company is looking at how much it pays in car insurance to people of different ages. It records the results as follows:-

		Average claim (£) 0-500	501-1000	1001-2000	2000+
	18-25	0	3	5	8
Age	26-35	3	7	12	4
	36-55	5	9	2	0
	56+	8	4	1	1

BDIIIA
BDIIID What does this tell the insurance company about the amount of money it is likely to pay in claims as it's clients get older?

. . . .
. . . .
. . . .

10. Two teachers both teach the same class. One teaches Maths the other Science. They both give the class a test in their subject. They decide to see if people who are good at Maths are also good at Science. The results were:-

	Anne	Brad	Chen	Dave	Erin	Fred	Gopi	Hana
Maths (%)	78	54	33	35	68	77	55	62
	○	○	○	○	○	○	○	○
Science (%)	33	36	41	22	43	35	41	39
	○	○	○	○	○	○	○	○

BDIB (a) Because the marks in the Science test are not so good the teachers decide to look at the positions. Write 1 under the person who came top, 2 under the one who came second and so on. Do this for both tests.

BDIIK (b) What does this tell the teacher about the marks in the two tests?

☐ Pupils who are good at Science are usually good at Maths as well

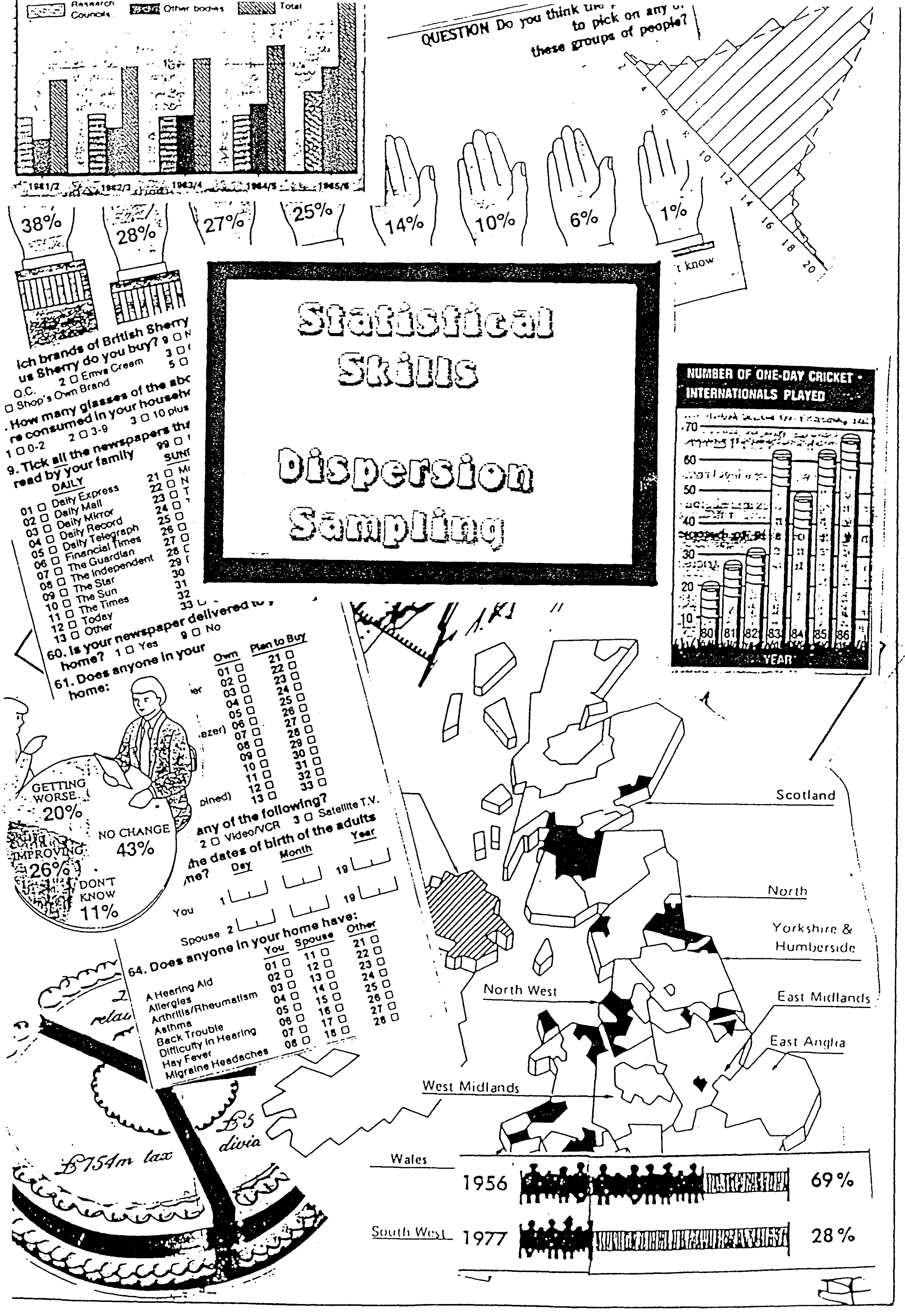
☐ Pupils who are good at Science are usually bad at Maths

☐ Pupils who are good at Science could be good or bad at Maths

END OF TEST

Statistical Skills

Dispersion Sampling



ANSWER AS MANY QUESTIONS AS YOU CAN - DO NOT WORRY IF THERE ARE SOME QUESTIONS YOU CANNOT ANSWER.

THIS TEST IS DESIGNED TO FIND OUT WHAT PEOPLE OF YOUR AGE KNOW

TAKE CARE WITH QUESTIONS YOU FIND EASY

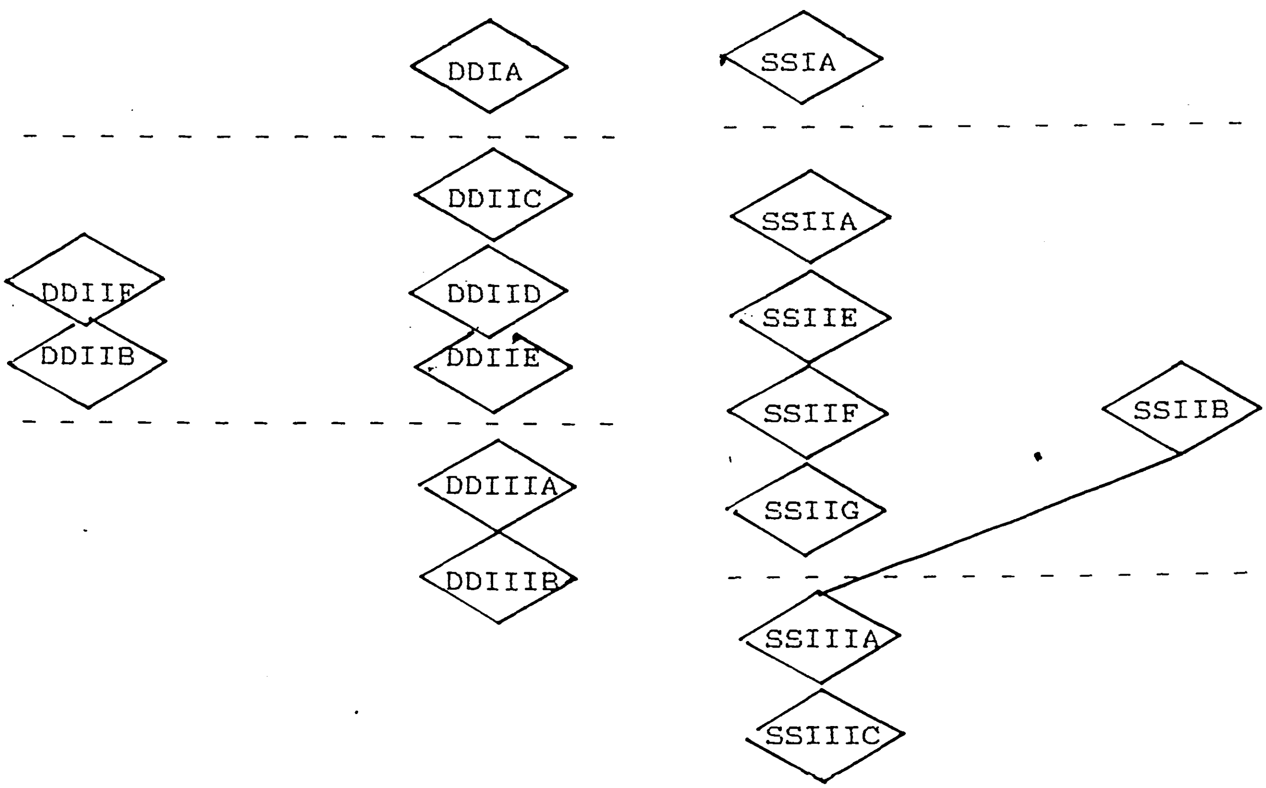
THERE IS NO TIME LIMIT BUT YOU SHOULD BE ABLE TO FINISH IN ONE HOUR

PERSONAL INFORMATION

DATE OF BIRTH ....

AGE ON 1ST JAN 1991 ....

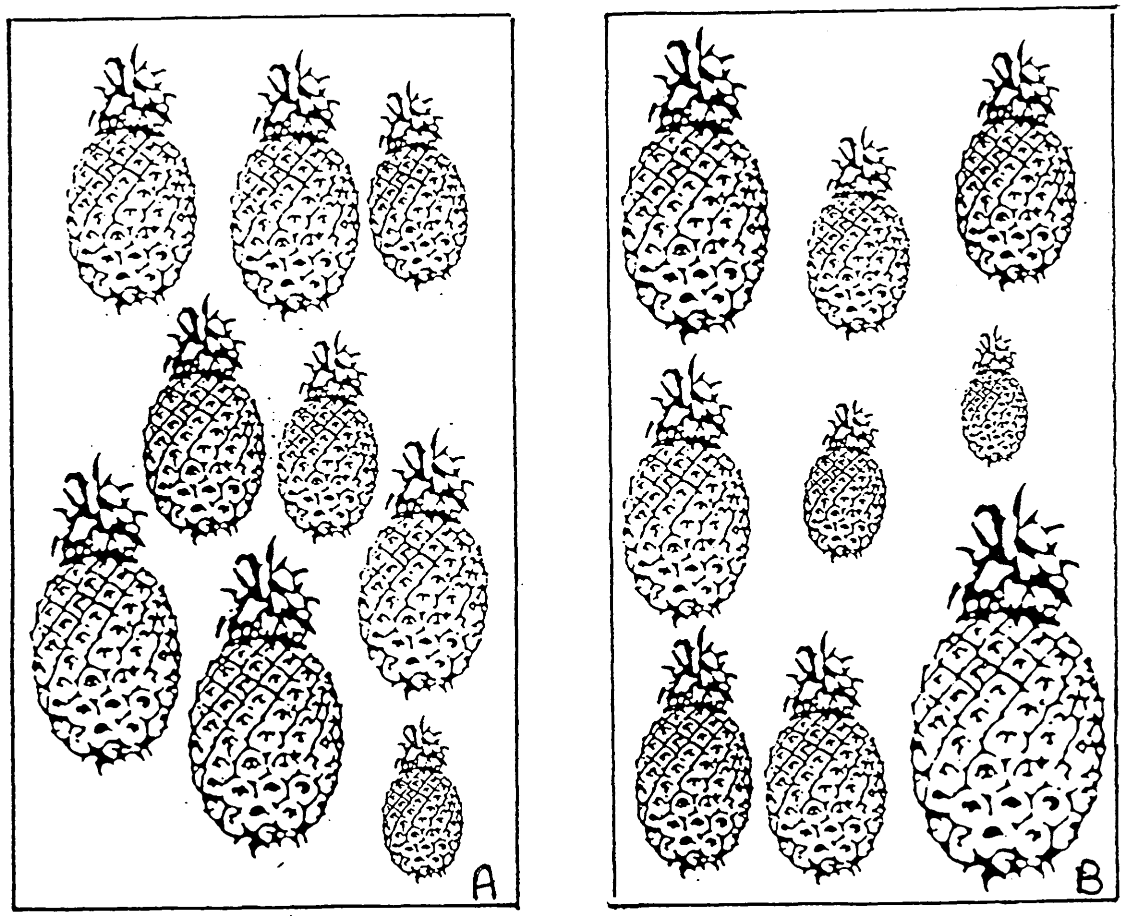
ANSWER GRID - DO NOT WRITE ON THIS



Measures of Dispersion

DDIA 1. A greengrocer goes to the market and is offered a choice of boxes of pineapples. These are shown in Data M. He notices that the average size of pineapple is the same in both boxes. What is different about the boxes of pineapples?

....
....
....
....



DATA M

DDIIB 2. A scientist claimed that boys heights are less varied than girls. A class of ten children measure themselves. Their heights in cm are:-

BOYS: 135, 142, 140, 139, 137

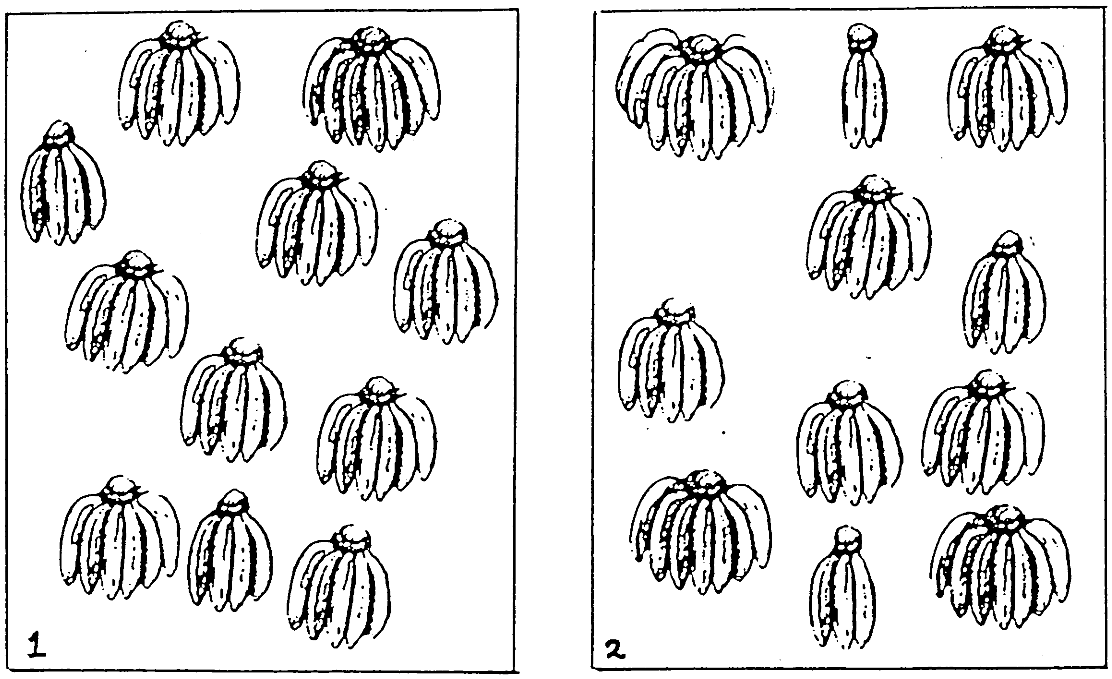
GIRLS: 128, 131, 138, 129, 134

Say whether you think the scientist is correct and give her your reason.

....
....
....

DDIA 3. At the same market they are selling bunches of bananas. In Data N are two boxes of bananas with the same average number of bananas per bunch. What is different about the number of bananas per bunch in the two boxes?

....
..
....
....



DATA N

DDIIF 4. A group of children measured their waists in cm. They listed these as:-

45, 56, ... 53

Someone calculated that the range of measurements was 11cm.
It was then discovered that the tape measure had the first 5 cm missing. How does this affect the range of waists?

....
....

5. Each pupil went round taking different measurements on others in the class. Some wrote them down in order, others did not. Find the range of values in each case.

DDIIC Head Circum. 36, 36, 38, 39, 39, 41 Range=.....

DDIIC Handspan 12.1, 12.3, 12.6, 12.8, 12.9 Range=.....

DDIID Shoe-size 5, 8, 4, 9, 7, 6, 6, 7 Range=......

DDIID Armspan 97, 88, 76, 87, 103, 85, 90 Range=......

DDIIF 6. Some children have grown sunflowers. In Data O the first picture shows what they looked like after one week.

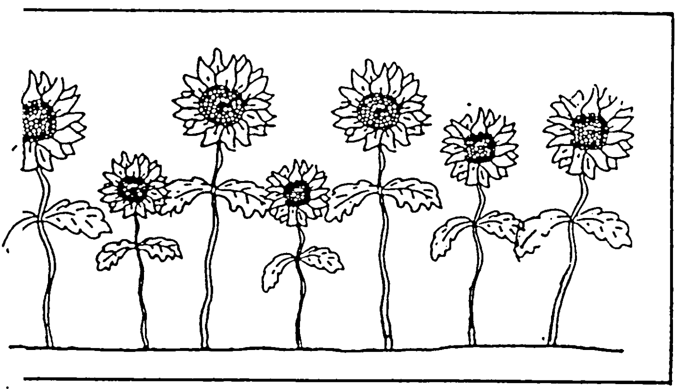
Amit said that there seemed to be a lot of difference in how much they had grown.

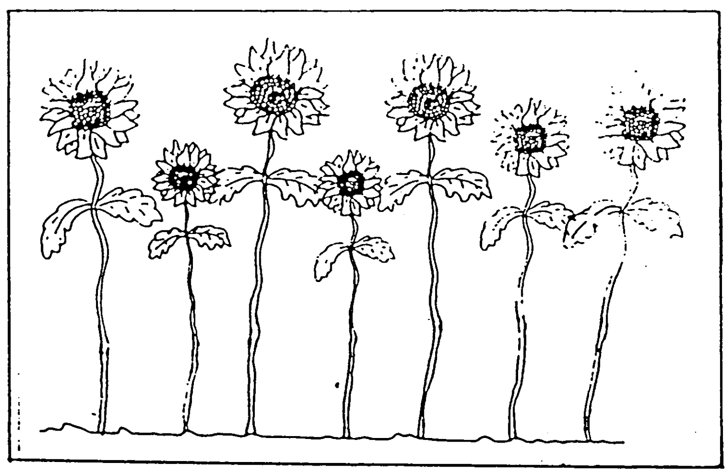
After two weeks they had all grown exactly 20cm and this is shown in picture 2.
Which of the following statements do you agree with most about the heights in the two pictures:-

☐ The heights are now more spread out after 2 weeks

☐ The heights have the same amount of spread

☐ The heights are less spread out in the second picture





ATA O

DDIIE 7. A brother and sister were arguing whose exam grades were more spread out. Emma said hers were and Damien said his were. Their grades were:-

Emma E, D, E, E, D, D, A

Damien D, E, B, E, C, B, E.

Which one of the following do you agree with most (tick the box):-

☐ Emma's grades are more spread out as she went from A to E

☐ Damien's are more spread out as he had 4 different grades Emma only had 3

☐ Emma's grades are less spread out as apart from the one A grade the rest were D's and E's.

DDIIE 8. A paper boy was collecting his Christmas Box. The amounts given people in one street were:-

30p, 50p, 50p, 40, 50, £1, 50p, £1, 50p, £1, £5, 50p

Which of these statements best reflects the differences people gave him?

☐ There was a lot of difference in what people gave as one person gave 30p and another £5

☐ There was not much difference in what people gave as most gave 50p or $1

☐ Therewas quite a bit of difference in what people gave as nearly half gave 50p and nearly half gave a pound.

9. The lower quartile is the number which is a quarter of the way up a set of data when it is put in order. The upper quartile is three quarters of the way up.
For each of these sets of data below give the quartiles:-(use the space as you need it)

		LOWER QUARTILE	UPPER QUARTILE
DDIIIA	2,3,5,7,8,9,13,15,17,21,25	.....	.....
DDIIIA	54,47,33,76,87,56,48	.....	.....

DDIIIB

Mark	Number
1-5	4
6-10	8
11-15	6
16-20	5

LOWER QUARTILE ..... UPPER QUARTILE .....

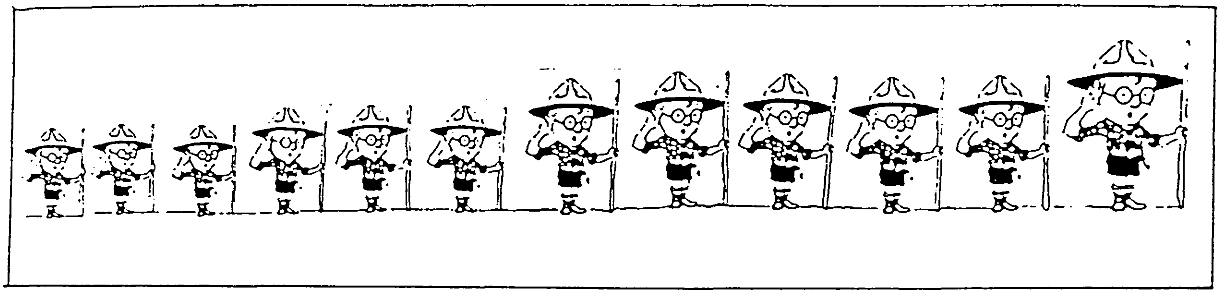
DDIIIB

Length	Number
2.0-2.9	2
3.0-3.9	7
4.0-4.9	4
5.0-5.9	2

LOWER QUARTILE ..... UPPER QUARTILE .....

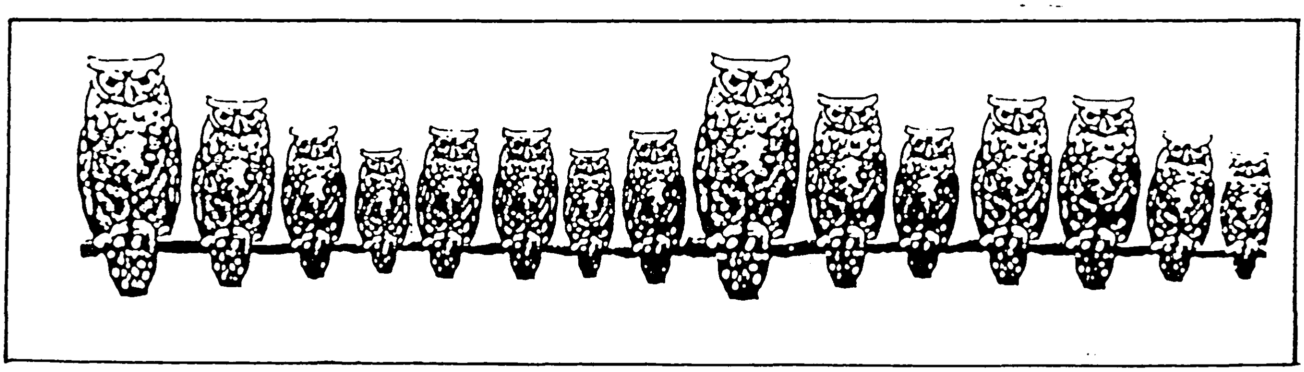
Sampling Skills

SSIA 1. A scoutmaster asks his troop to line up in order of size. He wants to select four scouts evenly selected from the troop. Put a cross under the four you think he should pick.

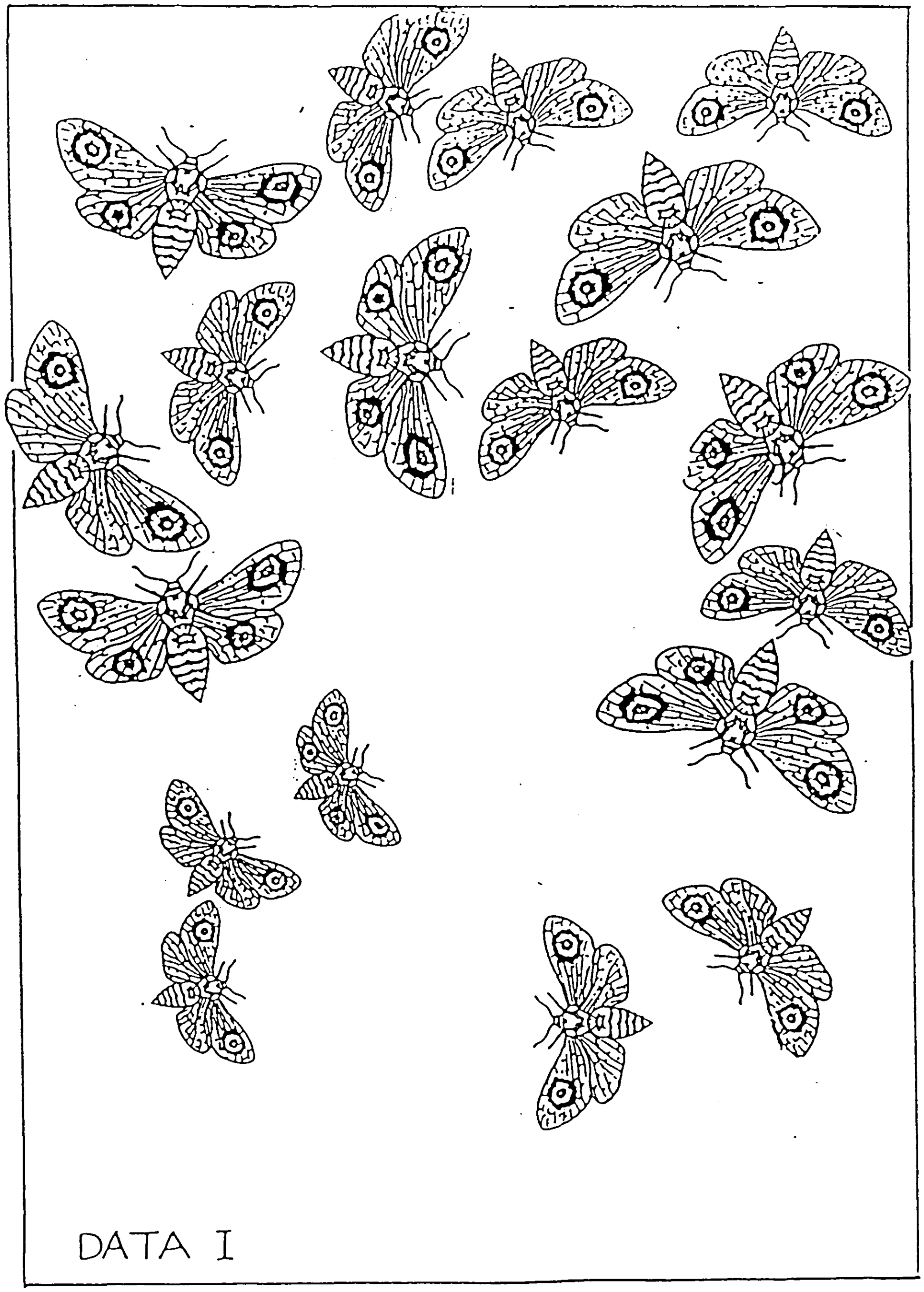


SSIA 2. Some owls are sitting on a branch. A scientist wants to select a range of sizes by going along the line and picking out at regular intervals. Put a tick under the five you would choose.

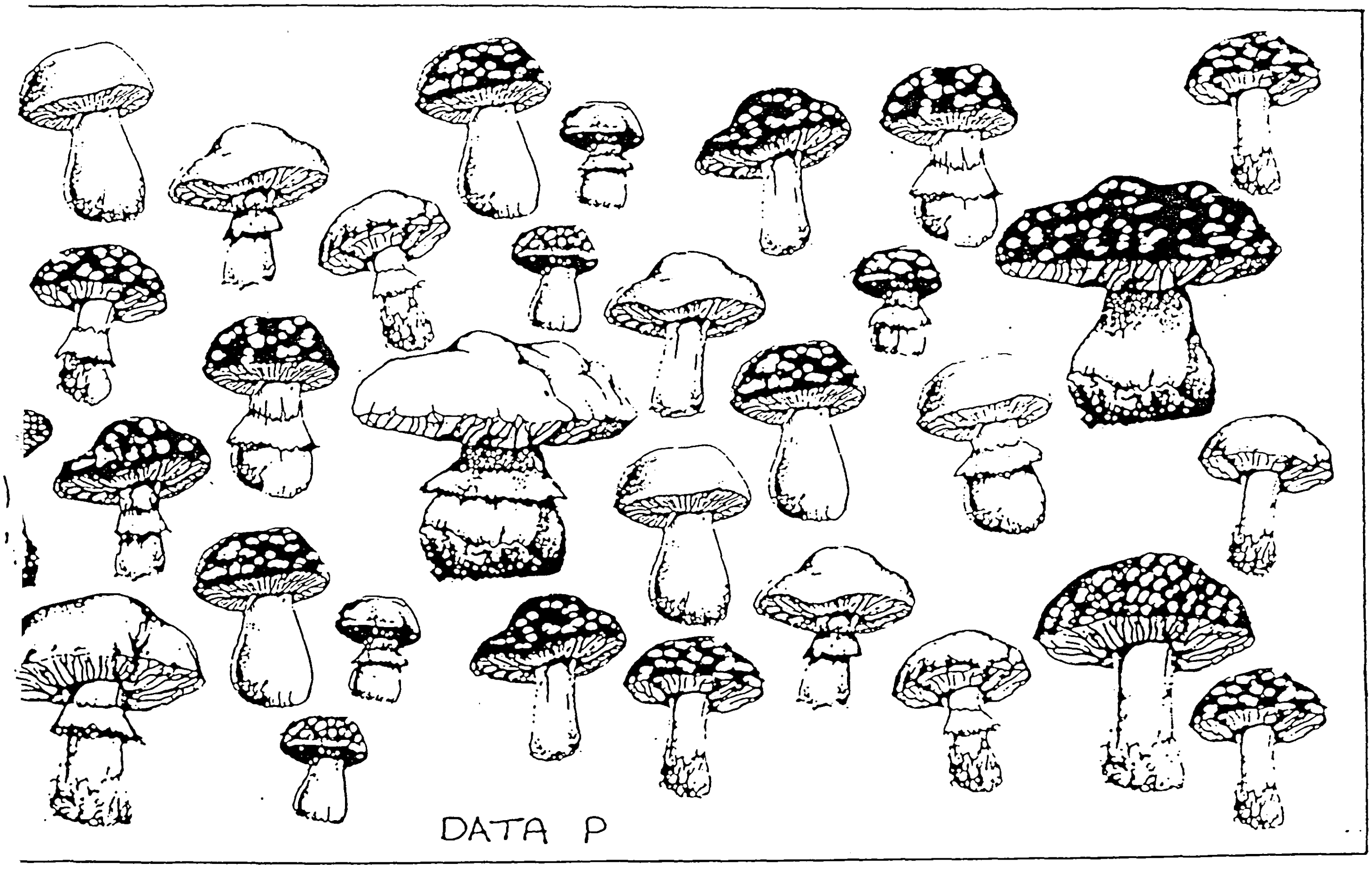
Do you think this is a fair method?...........
....
....
....



SSIIA 3. In DATA I are some moths which have one or two spots. An insect collector wants to select 6 moths which are typical of all the moths. He notices that there are more 2 spotters than 4 and wants to reflect this in the 6 that he chooses. Circle the six you would choose.



SSIIA 4. A toadstool expert collects two types. One type has spotty tops, the other plain. Data P contains 30 toadstools, 18 spotty and 12 plain. For further research she wants to keep 5 toadstools which have the same balance of spotty/plain as the whole lot. Put a circle round those you think she should keep.



5. An explorer in the Amazon discovers a new beetle which she calls a 'love bug'. She brings back 100 specimens and takes a picture of them. This is shown in DATA D2. She wants to look at the number of spots these beetles have. Unfortunately ink has been spilt on the only picture!

(a) Pick any 10 beetles which have roughly the right proportions of spots and put a ring round them. How many of your 10 have 2 spots?
. . . .

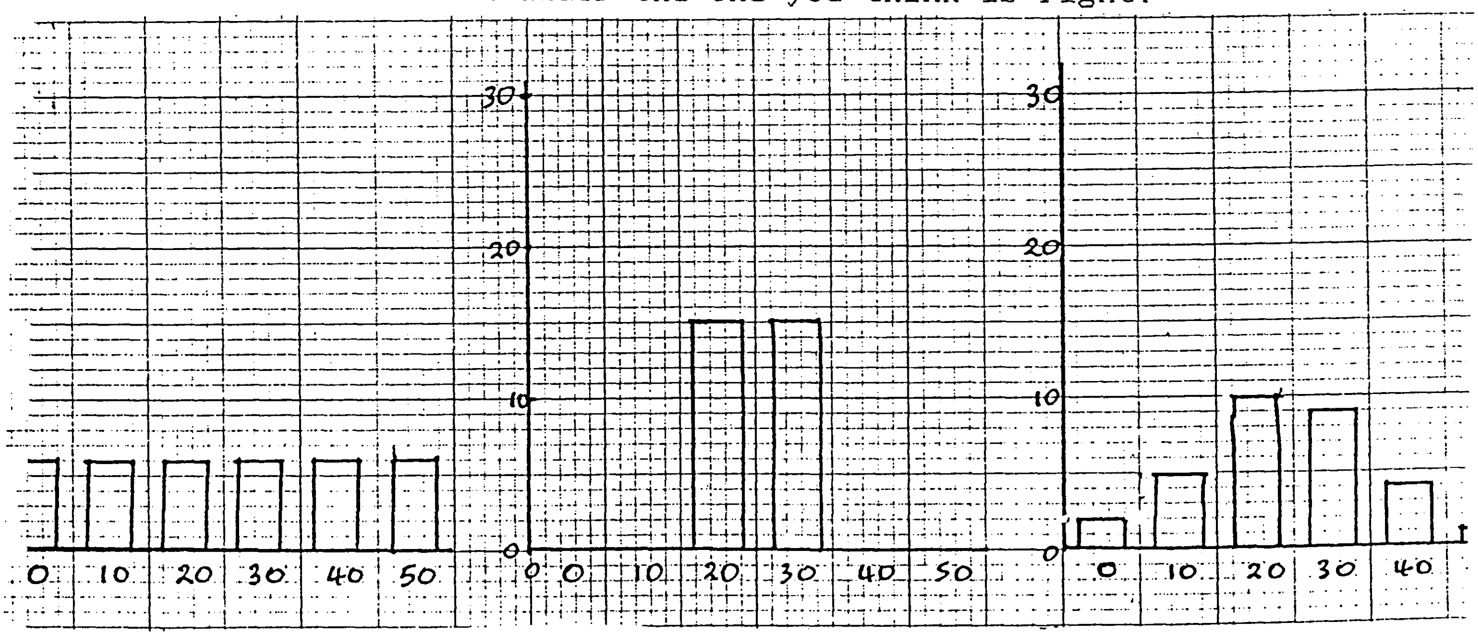
SSIIE (b) From your sample of 10 estimate how many 2 spotted love bugs there are in the whole 100. Show how you got your answer.
. . . .
. . . .
. . . .

SSIIF (c) How close do you think your answer is to the true answer(Tick the box)

☐ Exactly right ☐ Fairly close ☐ Nowhere near

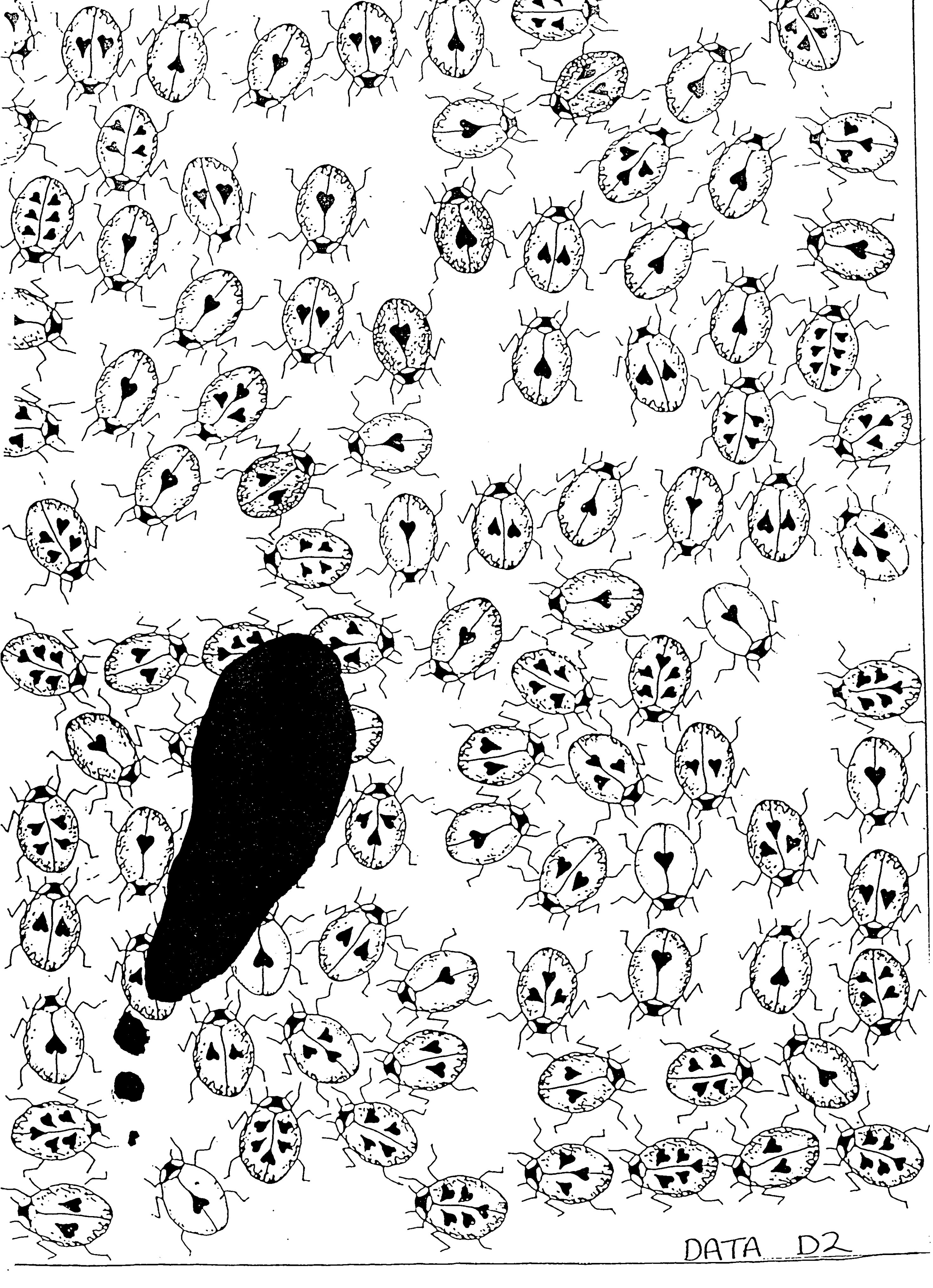
SSIIG How could she get a more accurate estimate without counting them all?
. . . .
. . . .
. . . .

SSIIIA (d) A class of 30 children all did the same thing to 'guess' the number of 2 spotters. The teacher put all the results on a graph. Only one of these graphs is the right one. Tick in the box under the one you think is right.

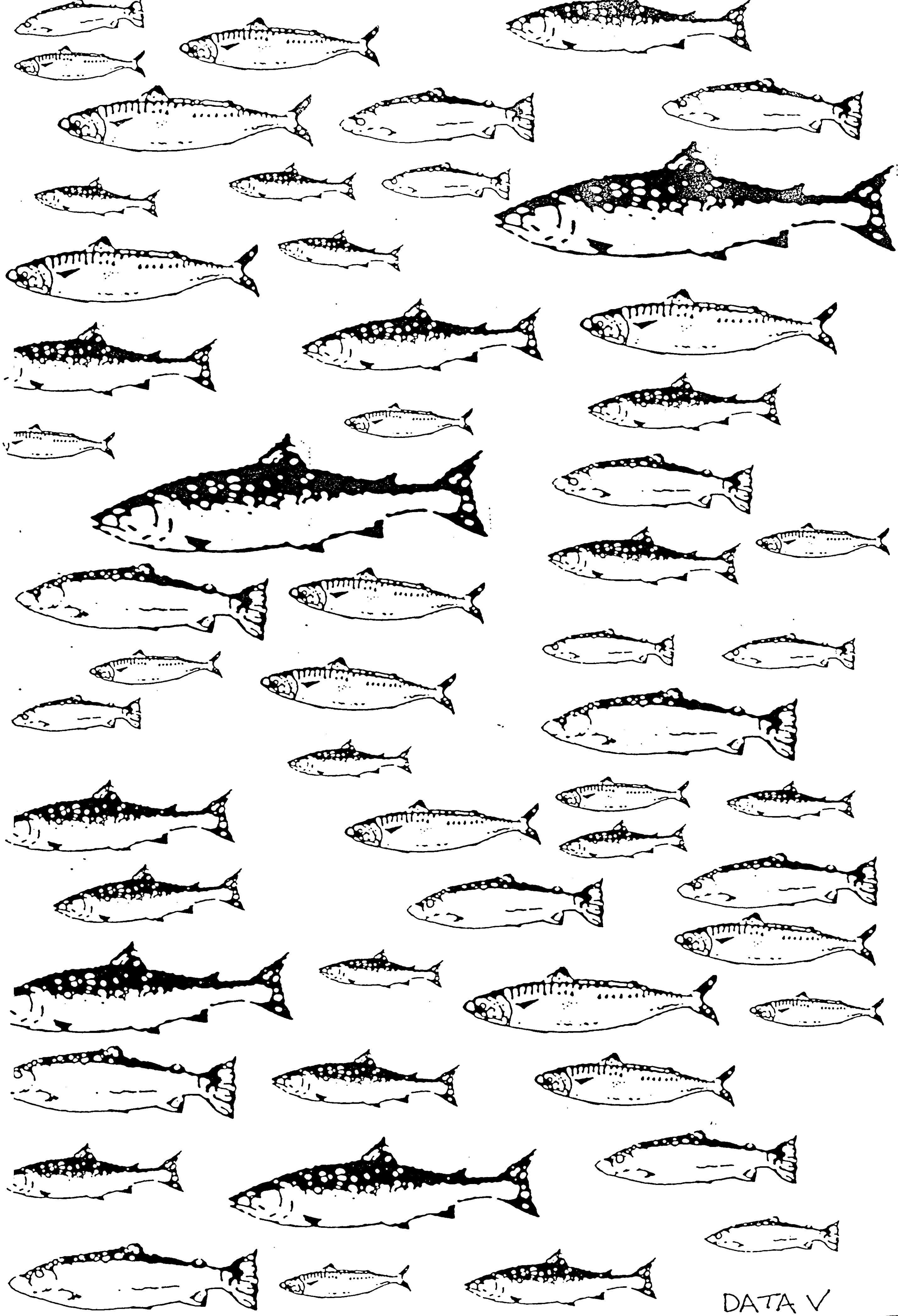


SSIIIC (e) Put a tick in the box for any of the values below which you think are likely to be the real answer:-

☐ 25 ☐ 30 ☐ 35 ☐ 40 ☐ 50 ☐ 20 ☐ 15



DATA D2



DATA V

a fisherman one day. He wants to find the average length of all his fish but cannot be bothered to measure them all. He decides to measure only five and calculate the average from this.

(a) Choose five fish which show a reasonably fair reflection of the different sizes and circle them.

SSIID (b) Measure the lengths of your 5 fish to the nearest 0.1 cm. Write the lengths below and find the average:-

A: B: C: D: E:

Average = cm

SSIIF (c) How close do you think your average length is to the average length of every single fish?

☐ Exactly right ☐ Fairly close ☐ Nowhere near

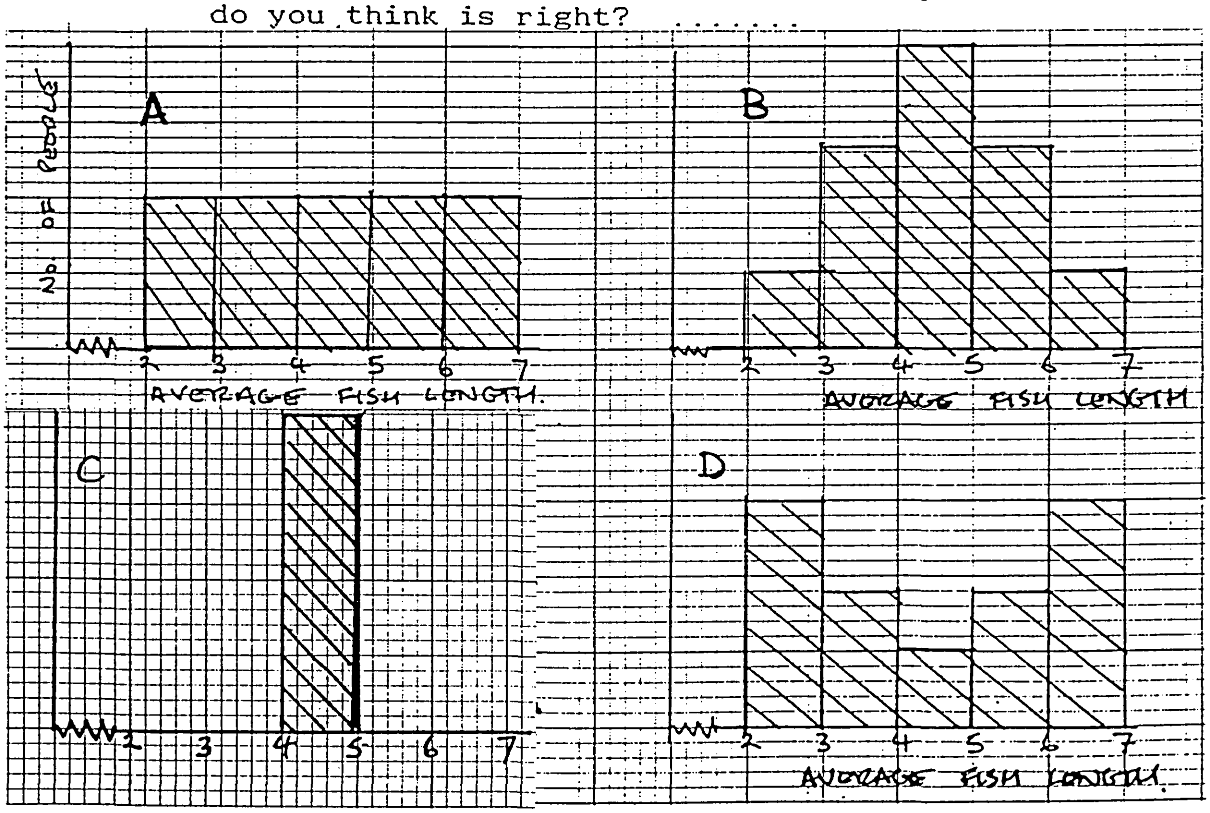
SSIIG How could the fisherman have got a more accurate guess at the average length of all his fish without measuring every one?

....

....

....

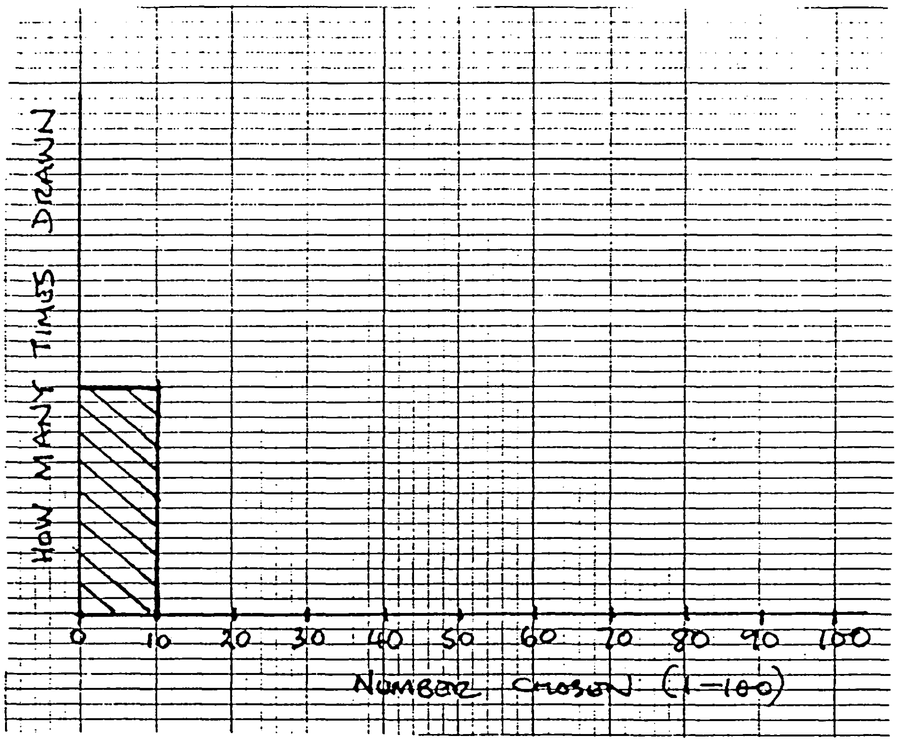
SSIIIA (d) The fisherman asked lots of his friends to pick five fish and work out the average length of fish. He shows them on a diagram. Only one of these is right the rest are made up. Which do you think is right? ....



SSIIIC (e) Which of these answers are likely to be the average length of all the fishes? (Tick as many boxes as you like)

☐ 4.4 ☐ 5.2 ☐ 5.9 ☐ 6.6 ☐ 7.7 ☐ 3.3

SSIIB 7. (a) A class has a bag full of bingo numbers (1-100). They take it in turns to pick out a number and put it back. The do this lots of times. On the graph below sketch what their results might look like.
The first block shows how many times numbers in the range 1-10 were picked.



SSIIB (b) They then measure how tall they all are in cm. Sketch the graph that you think they would have drawn. The tallest person is 184cm, the smallest 138cm. The block drawn shows the number of heights between 135 and 140cm.