CHILDREN'S UNDERSTANDING AND USE OF A DATA BASE

by

Maria José Bezanilla Albisua

A Thesis submitted in fulfilment of
the requirements for the degree of
Doctor of Philosophy

University of London
Institute of Education
April 1992

ABSTRACT

The focus of the present research is on children's understanding and use of data bases.

Previous research in this area is limited and has been mainly machine-centered. Some studies have been done to analyse the effectiveness of data bases compared to more traditional methods (paper and pencil) in information handling activities; and some others have focused on the comparison among several data base systems in order to see to what extent the system characteristics (data base structures - command/menu driven systems) influence the process of information retrieval. However very little attention has been paid to the process of teaching and learning data bases, to the nature and structure of data base tasks, and the factors that affect students' performance on such data base tasks. This latter perspective, task-centered, is the one taken by the present research.

The present research investigates the possible existence of a theoretically proposed five-level hierarchy of data base tasks. 13 and 16 year old students' performance on two different types of data base tasks using GRASS is examined. The research also involves questions about the skills and knowledge that students possess and which affect the quality of data base interrogation activity, such as understanding of the nature of data bases, understanding of the nature of data base search and structure, and understanding of Boolean logic in the phrasing of complex queries. Children were tested on a range of such tasks before and after being trained in the use of GRASS.

The results show that children at 13 and 16 can perform simple data base operations and tasks rather successfully. There is some hierarchy in the difficulty of tasks, but not precisely that proposed theoretically. Tasks involving two variables, and relations between them, are relatively difficult, particularly for the younger group. Tasks requiring simple retrieval and sorting data are the easiest. Logical operations present difficulties, especially in the construction of logical sentences, with simple AND combinations being the easiest. OR is sometimes treated as AND, particularly when interpreting negative OR propositions. Contradictions and tautologies are very difficult both to interpret and construct. Children seem to have an adequate if simple model of a data base in terms of the types of information a data base can and can not contain. However, their models of the ways information is kept, organised and found in a data base are not so adequate.

To my parents

ACKNOWLEDGMENTS

I am especially grateful to Prof. Jon Ogborn, my supervisor, who has provided me with expert advice and encouragement from the very first step to the completion of this thesis. His kind, continuous and critical supervision, active support and help have been invaluable.

I would like to thank Mr. Leslie Beckett for his interest in all phases of this work, as well as for his help in the school selection process, and also for his critical review of the final draft of the thesis.

I would also like to thank Prof. Jim Eggleston for his interesting suggestions on the methodological aspects of the research.

My thanks also go to Dr. Richard Cowan for his advice and help with respect to the Logic aspects of this work, and to Dr. Harvey Mellar for his reading suggestions and valuable discussions throughout. My thanks go to Mr. Dick Booham too for his comments on the final draft of my thesis.

I am indebted to the students who have participated in the various tasks of this research, as well as to their teachers: Mrs Sheila Morris, Mr. Malcolm Frazer and Mr. Ian Mulelly for their constant support and encouragement during the data collection phase.

The emotional support which I have received from friends and family during this research has been inmense. I am especially grateful to my friends Gloria Gutierrez and Padma Anagol for their unconditional friendship and stimulating discussions, and to my friend John Barrett for his optimism, support and good advice. Thanks are also extended to the research students and staff of the Science Education Department.

Finally, I would like to express my most sincere thanks to my parents, brother and sisters for their financial and moral support during the three years and a half I have spent in London. This work would not have been possible without their unfailing support, concern and encouragement.

CONTENTS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

Chapter 1 INTRODUCTION

1.1 The problem

It is more than a decade since computer tools (e.g. data bases, word processors and spreadsheets), first became available in forms appropriate for schools, and started to gradually be used widely by teachers and pupils in all curriculum areas. Before then, drill and practice programs or tutorial modes of courseware were the types of software most commonly used in schools. These applications still play an important part in classroom activities, but they are gradually being replaced by the tool-like capabilities of the computer. Unlike tutorial programs, which could be used for only one task, these open tools appear to allow any number of tasks to be achieved and to provide opportunities to apply problem solving strategies to many areas of the curriculum. Besides, they accurately mirror the uses of the technology in the world of work.

The inevitable lack of expertise of most teachers in the use of these tools has led them to feel doubts about their own abilities to use the technology and about what to do with them in the classroom. The need for research into how best to view their use and into how to include them in everyday teaching became urgent. The present research focuses on one of these computer tools, the data base. Nowadays, because of increasingly available masses of data, data bases, associated with the handling of information, have acquired extreme importance. They are already being used in many primary and secondary schools, but research is needed to explore in more depth to what extent they are educationally valuable and how they can help information handling, as well as in what way they can become an integral and fundamental part of the curriculum.

Those who teach computer literacy courses that have a unit on data bases point out that it is typical for students to go through three stages in learning to use a data base.

First, they learn to work with the data in a data base created by someone else; they practice searching, sorting and displaying the data. Second, they learn to build a data base using a structure provided by someone else; they engage in entering and editing data. Finally, they learn to design the data base; they create a structure for organising some collection of information. The present research focuses on the first stage, the data base interrogation activity. It examines a variety of tasks to be done with an already created data base and some of the factors that influence the query formulation process.

There have been many claims about the benefits of using computer data bases in the classroom. Their potential for promoting the development of children's thought processes and their popular use in the "real world" are two such claims. The aim of this study is not to develop an argument for the promotion of data base use, but rather to get some empirical data about what children at a certain age can and can not do with a simple data base, and about the skills and knowledge children have and need to develop to use a data base effectively.

Examples of experiences of data base use in the classsroom context have been widely published. However, empirical research in this area is still very limited. Some preliminary studies have focused on the comparison between a data base and its non-computer counterpart, the card filing system. This methodology, however, has been criticized as neither useful nor convincing, on the grounds that it is difficult to explain how a card file might be a powerful tool for carrying out a process of substantive inquiry.

Other studies, in the line of research of Human-Computer Interaction, have focused on the comparison between different data base interfaces (menu-driven and command driven systems) and between different data base structures (tabular, hierarchical, relational, etc.) and their influence on the process of information retrieval. These researches are mostly machine-centered and, although useful for software design and evaluation purposes, provide little direct contribution for classroom practice.

The present research is not machine-centered, but is task-centered. It attempts to investigate children's understanding and use of data bases, with a particular emphasis on the nature and structure of the tasks. It involves defining tasks of different levels of difficulty, seeing whether they form a hierarchy, and finding which aspects of the tasks, related and non related to the computer, may present particular difficulties. It also involves the design of "real life" tasks and other more artificial tasks to see how children's performance may be affected by the particular characteristics of these two types of tasks, and what skills may be required for them. The line of research adopted agrees with the view taken by Olds (1986) when he states: *"The secret for understanding a data management program is to use the program to do a large number of tasks of different kinds or to solve a range of problems that require data analysis. The more the program is used to carry out meaningful tasks, the better the tool is known and understood. Therefore, a data management program is defined by what you can 'do' with it ".*

Once the the potential of the tool has been tapped and a wide range of tasks that can be done with it have been defined, other questions emerge. For instance, it is important to find out how the tasks will work in the classroom and how students will respond to them. Will the tasks be equally difficult for children of different ages? What are the difficulties of the various types of tasks? What are the factors related to the machine and related to the student, determining ease and difficulty? These are all interesting questions, which this research attempts to address.

Recently, some studies have looked at difficulties involved in the interrogation of a data base, and the factors that are important for the quality of queries. The results reveal that children need to understand the structure and the search method of the data base program to achieve successful outcomes. Children's understanding of Boolean logic is also reported as affecting query formulation activities. The present research also includes questions about children's logical skills and their knowledge about data bases. These can be stated as follows: How well can children understand the logical operators AND and OR? What is their knowledge relating to the general nature of data? What is their knowledge relating to the general nature of data base search and structure? Does their understanding of logic improve after having learned how to use a particular data base? Does their general understanding of the nature of data bases improve?

Factors that affect the query formulation process arise one at a time in different pieces of research, and the literature shows a lack of a synthetic view of them. This research will try to see how such factors relate, and how they affect performance on the proposed hierarchy of data base tasks in an attempt to make a more comprehensive study of children's use of data bases.

1.2 Methodological aspects of the research

The data base used in this research is GRASS. GRASS was chosen because it is currently one of the most widely used data bases in schools, and because of its simplicity and ease of learning and use. The data file chosen (POP) contains information about the population of the world. This subject is addressed at different stages of education, and it is hoped that children's familiarity with it will be sufficient so as not to interfere with the search process.

The nature of the experimental work is of two kinds. On the one hand, we have children's work on the computer when solving the various data base tasks, and on the other hand, we have children's responses to the four questionnaires designed to test their understanding of data bases and of logical operators. One of these questionnaires consists of open-ended questions and requires a qualitative methodology. Children were given these questionnaires before and after having been trained in the use of GRASS and having solved the data base tasks, so that possible improvement in their general understanding of data bases and in their logical competence could be detected. Performance on data base tasks was measured by looking not only at children's responses to the tasks, but also at the ways they reached their answers by interacting with the computer (search strategies).

Subjects of two different age groups participated in this research: 13 and 16 year olds. They had all used computers before, so they were familiar with the computer keyboard, but none had worked with any data base before. It is hoped that comparison of the performance of these two groups of students will help to establish possible performance benchmarks for certain data base tasks and to throw some light into the cognitive demands of using a data base, and possible influences of development or other learning.

1.3 The potential contribution of the research

It is hoped that this research will offer some contribution both to teachers and curriculum planners. Teachers could benefit from the results of this work by being able to know better what the knowledge and abilities regarding data base use of an average 13 or 16 year old may be, and thus, being helped to adapt their teaching strategies according to this knowledge and ability.

The present research is novel in attempting to develop a wide range of data base tasks of different hypothetical levels of difficulty, and of different types (e.g. "real life" tasks). These could be used by teachers for diagnostic purposes. At the least, the tasks presented in this work could serve as an example of the multiple possibilities that a data base offers and could suggest ideas for the design of further tasks. At the most, the results of this research could offer some guidance to teachers as how to meet certain National Curriculum goals.

It is also hoped that the results of the present work will influence curriculum development positively. This thesis presents some empirical data about what children at different ages can and can not do with a data base. Consideration of these results may favour the establishment of maybe more realistic and specific attainment targets regarding data base use for the different key stages, and shed some light on the issue of the evaluation of classroom outcomes following the use of data bases. It may also clarify the problem of measuring the individual progress of children.

1.4 The organisation of the thesis

Chapter 2 is a review of the research in the area of children's understanding and use of data bases, with particular emphasis on data base interrogation activity. The "logical" hierarchy of data base tasks, which is to be used in the rest of the thesis, is described. Also, there is a short section on data bases and the National Curriculum.

Chapter 3 outlines the questions this research attempts to answer. It also describes the tasks and instruments of data collection used to approach them, and the rationale for the tasks and instruments.

Chapter 4 presents the results of a small pilot study on the viability of the tasks and questionnaires designed, and contains methodological considerations related to modifications made to the instruments of data collection to be used in the main study. It also describes the sample and the experimental design of the research, as well as the training and administrative details of the tasks.

Chapter 5 discusses the results of children's performance on the two types of data base tasks used ("Exercises" and "Report editing" tasks), and their relation to the "logical" hierarchy of tasks.

Chapter 6 presents the results of the tasks concerning logical operations, and discusses children's difficulties in understanding logical operators, as well as in interpreting and constructing logical sentences.

Chapter 7 discusses the results of a short questionnaire which investigates children's understanding of the nature of data, or the types of information that can and can not be retrieved from a data base.

Chapter 8 investigates children's understanding of the nature of data search and structure for each of the tasks defined in the "logical" hierarchy of tasks. The results of an open-ended questionnaire designed for this purpose are presented.

Chapter 9 attempts to summarise the main findings of the research and presents suggestions and advice for teachers and curriculum planners.

Chapter 2 LITERATURE REVIEW: CHILDREN'S UNDERSTANDING AND USE OF DATA BASES

2.1 Computers in Education

Computing has rapidly become an essential part of education whose influence reaches across the normal curriculum boundaries. The decreasing size and cost of microcomputers during the last twenty years has led to a bigger use of computers and related information technologies in society. Nowadays, with an average of one computer for every 67 pupils in primary schools and one for every 32 in secondary schools (1988 DES Survey - Boyd-Barrett, 1990), the issue of how to use them in a proper way as an integral part of the existing curriculum has become extremely important.

When the first computers were introduced in schools, in the 1960s, the computer was initially the object of study in its own right, either through awareness and literacy courses or through computer studies courses, in which the teaching of programming languages was the main activity. There were two main arguments in favour of such studies. The first one concerns the needs of society. Since computers play an important role in peoples' everyday lives, students should be taught their use in preparation for their later professional careers. The second argument concerns the teaching of programming. Here it was asserted that learning to program is an ideal medium for the development of general problem-solving skills, and for the exploration of other curricular areas, such as Mathematics.

Since then, from the late 1970s onwards, there has been a huge growth and improvement in the nature of software available and in the activities which can be supported with it. Indubitably, the key to the current use as well as future potential of the micro lies in the software which is used. Different kinds of software require different abilities and promote also different ones, and the choice to use one or another type of software presupposes a particular educational philosophy.

Software has been classified in a number of ways. Perhaps the most sensible classification is that which distinguishes those programs which are an end in themselves, in the sense that they have a well defined goal, and those open tools, with which any number of different goals can be achieved. Among the first group we can find CAL and drill-and-practice programs and some simulations. Among the second, we can find word processors, data bases, spreadsheets, and programming languages. According to Taylor's (1980) classification of computer applications, the first group of software would correspond to the computer functioning as a "tutor" where the computer is basically programmed to present information, followed by questions, and then feedback to students about their responses. The second group of software would correspond to the computer functioning as a "tool", as a medium for accomplishing some task, or as a "tutee", where the student tutors the computer by using a programming language.

Nowadays there is a growing emphasis on having children use general purpose software tools, and there are two main reasons for doing so. First, in the "real world" computers are almost exclusively used as tools; second, these open tools support the problem solving approach to education against the content-oriented one, and the idea that education is not only about "knowing what", but "knowing how" as well. It is also said that the use of the computer as a tool allows children's minds to develop through the exploration of computer-simulated "microworlds" and encourages the development of higher level cognitive skills (e.g. Papert, 1980).

In the following sections we will concentrate on one of these computer tools, the data base, in an attempt to see to what extent it is an educationally valuable tool, as well as in what ways it can help information handling, and become an integral and fundamental part of the curriculum.

2.2 What a data base is

In a broad sense, a data base is just a collection of data, usually organised in a particular way. Card files, address books, and file cabinets for example are all quite useful and common non-computer methods for organising data bases. With the advent of the computer, however, came a type and level of data base organisation never possible before, which enables the user to enter, store, update, access and manipulate information with great flexibility. These computer data bases are what we simply call now data bases. Data base management systems and information retrieval systems are also common terms used to describe computer data bases.

Date (1990) in his book *An Introduction to Database Systems* defines a data base as follows:

> *"A database system is basically a computerised record-keeping system; that is, it is a computerised system whose overall purpose is to maintain information and to make that information available on demand" (Date, 1990, page 5)*

This definition introduces the two main elements of any data base: an organised body of information or data, and a program or set of instructions which allows the user to store, update, retrieve, organise, sort and analyse that information. Also, the most flexible data bases include the facility to display data in a variety of graphic forms, encouraging a more meaningful interpretation of data.

Although all information stores, including books, directories, and data bases are displays of organised data, the structure of that organisation is clearer in computer data bases. Also, as the information is kept on a computer, any item or combination of items can be recalled immediately. Basically there are four main ways in which data bases differ from other means of information storage, all of which facilitate information handling:

1. by being able to store a large amount of information
2. by providing a structure or framework for the information, so that it can be better organised
3. by being able to retrieve the information required more quickly
4. by being able to display the information easily in a variety of forms, such as tables, diagrams, etc.

These four big advantages, CAPACITY, STRUCTURE, SPEED and FLEXIBILITY, are what make data bases of great value. Although originally they were not specifically developed for instructional use, they are likely to offer remarkable opportunities in the field of education.

Data bases range widely in complexity, power and ease of use, the nature of the organisational structure of data, and the nature of data search and retrieval being the two key ways in which data bases differ one from each other. These will now be discussed in more detail.

2.2.1 The nature of data structure

The structure underlying most currently available data bases is usually categorised as hierarchical, tabular, network or relational. However educational data bases are mostly tabular, like GRASS, QUEST, and FACTFILE, or hierarchical such as SEEK and ANIMAL, all of which are currently in use in schools.

A tabular structured data base contains a collection of information about a particular topic called a *file*. The file is organised in *records*, each of which contains different headings or categories of information called *fields*. So, depending on the number of records and fields, we can find data bases of very different sizes. For example, if we think of a body of information about characteristics of different countries of the world, then the names of the countries, continents they belong to, population, etc. will constitute a file. The particular information specific to one country will be equivalent to a record, and the categories of data that will occur for each country are the fields. File, record and field are then three basic features of a data base. A data base can contain one or several files, but combination of information from more than one file is impossible with a tabular structured data base. One must work with one file at a time.

The relational data base is a step up from the tabular/matrix one. Its organisation is similar, records subdivided into fields, but in addition it will allow file by file searching and combining records from different files. The records searched from different files must have at least a field in common. Relational data bases are not commonly used in education yet, but the need for them is apparent. Underwood and Underwood (1990) found examples where children could be better served by a relational data base, even at primary level.

The hierarchical structured data bases organise data into records only. There are no fields. Data is represented in the form of a set of tree structures, and the operators provided for manipulating such structures include operators for traversing hierarchical paths up and down the trees. Interconnection of files is not dependent on commonality of field, but on a predetermined hierarchy established by the user upon setting up the system. They are used in education, but are often considered more rigid and inflexible than the tabular one (but see below). SEEK and ANIMAL are two of the most common hierarchical data bases currently used in primary schools.

A network data base is similar to a hierarchical one. Data is represented in a network, which means that each item of information should be coherently interrelated with all of the other items. The user has the options of interrelating records with other records or files with files. They are said to be comparatively difficult to design and maintain, and also very inflexible and difficult to extend (Burnard, 1987). Network structures with their pathways and common links and nodes are not used in educational systems yet.

So, unlike the one dimensional organisational structure of a book, data files have one ofa number of two-dimensional structures based on hierarchical, tabular or network organisations. Students may not be equally familiar with all types of organisation, and this might in turn affect efficient data retrieval. This problem has been addressed by a number of investigations of how people organise and retrieve data (Durding, Becker and Gould, 1977; Brosey and Shneiderman, 1978; Ray, 1985; Underwood, 1986; Underwood and Underwood, 1987). Durding et al (1977) confirmed that adults, when faced with word sets exhibiting a range of pre-defined organisational structures, were able to recognise and make explicit those structures. There was, however, a ranking of ease of use: lists, hierarchies, networks and tables. Rather than asking their subjects to construct organisational structures, Broshey and Shneiderman (1978) asked them to access information from data bases, with either a tree (hierarchical) structure or a tabular structure (relational). The tree structure proved to be an easier retrieval format, and it was also easier to commit to memory and reproduce, in comparison with the tabular structure. The two sets of experiments together suggest that after lists, the most accessible form of data is in hierarchies or trees. Other organisational structures may be used with some measure of success, however, if the student is made aware of the relevance of that structure to the task.

In contrast with the above results, a study by Ray (1985) on the effect of different data models on casual users' performance in writing data base queries shows that subjects using the relational model performed significantly better than those using the hierarchical model when writing the specification and navigational portions of the queries. (In the specification portion, the user must specify all attributes to be retrieved from the data base. In the navigation portion, the user must define the navigational path to be used in the information retrieval process). Ray suggests that the contrasting results are probably caused by the different dependent variables (memorisation and comprehension versus query writing). This is an indicator that additional research is needed to measure the correlation between ease of use and actual query writing.

Underwood and Underwood (1987) in an experiment with 9-11 year olds found that children were not very successful in discovering the inherent organisation of information and showed a tendency to impose a standard list structure on all data regardless of their inherent organisation. Hierarchies and tables were perceived as equally difficult by these young children, while networks were somewhat easier to construct. Another experiment showed that the children were able to extract data from existing hierarchies, networks and tables. A study by Underwood (1986) on primary children's classificatory abilities also confirmed there was no significant difference between the two methods of organisation tested, trees and tables.

Essentially these experiments suggest that data structures must be made transparent to children if they are successfully to organise or extract information from a particular structure, and that structures differ in difficulty of use and understanding.

2.2.2 The nature of data search

It is claimed that as well as the nature of the data structure of a data base, the methods of data search have also a big impact on the ease of use of the program and on the cognitive experiences presented to the child (Underwood and Underwood, 1987, 1990; Spavold, 1989; Brazier and Beishuizen, 1988; Freeman and Levett, 1986). Access to a data file is either through a menu or through the use of a command language. Menu-driven programs create questions for the user to answer, which may require "yes" or "no" responses or require reference to the contents of the file. The system repeatedly presents the user with list of options, from which one may be chosen. Execution of an action entails a fairly fixed sequence of option choices. The sequence of menu presentation is defined by the system. The user makes the choices. Command-driven programs employ a user or non-user vocabulary and a non-user defined syntax for the user to construct commands (questions) for the data base. Menu-driven programs are easier to operate but lack the efficiency and flexibility provided by the command-driven programs. Usually, preference is expressed in favour of menu-driven ones for the naive user and of command-driven ones for the experienced one, since the latter require a high degree of competence with the command language, sometimes quite complex, and a clear mental map of the data structure (Underwood and Underwood, 1990; Freeman and Levett, 1986; Marchionini and Shneiderman, 1990).

> *"Low frequency users may develop accurate mental models for what the system can do, but forget the details of system use. These users need menu and on-line reference aids. Frequent users, on the other hand, may prefer commands to expedite their use of the system" (Marchionini and Shneiderman, 1990, page 253)*

Brazier and Beishuizen (1988) did research in order to find out to what extent learning to work with a particular data base affects the ease with which the system is mastered, the strategies employed for the retrieval of information, and the user's mental model formed. The two educational data bases analysed for that purpose were GRASS, menu-driven, and QUEST, a command-driven system, and the examination of the results favoured the use of QUEST on all accounts. They attributed their findings to the structure of the system, in the sense that the commands are used as a basis for the analysis of the task to solve, thus structuring the problem solving domain, whereas sequences of menu options are less clearly remembered and comprehended by the students.

"The structure of the interface determines both the quantity and quality of the interaction. The user should be able to determine the sequence of actions within a transparent, well-structured environment, according to his/her analysis of the problem" (Brazier and Beishuizen, 1988, page 17)

This problem of transparency of the data structure was also addressed by Fitter (1979) when he suggested that in order for users to feel in control of the system they need to have an adequate knowledge of it, which in turn means that they must know where they have been, where they are, and where they can go from there. Essentially the user requires a mental map of the system but also a reason for navigating through the system.

Recently, interest is being shown in the design of data bases which use "plain English" for the formulation of queries. Natural language query techniques are an attempt to build a bridge between the ease and supportive nature of menu-drive programs and the flexibility of command-driven ones. They are an extension of the command system with the advantage that they will accept instructions phrased in English rather than the highly specified command language. Shneiderman (1980), however, lists a number of problems that natural language systems need to overcome. The unrealistic expectations of the computer's power and the ambiguities of the English syntax are two of them. As with the relational data base, natural language techniques have had almost no impact on education up to now.

2.3 Data base skills

As has been said previously, data bases are powerful content-free programs that enable children to handle data by storing, retrieving, organising and analysing information in a very efficient way. These general information handling skills can be gathered under the three main stages involved in the use of any data base: the design, the interrogation and the interpretation of the data file. These activities in their turn involve a number of skills, like gathering information, organising and entering data into the computer, formulating a search strategy and dealing with the result of a search. Each stage (creation, interrogation and interpretation) constrains subsequent stages. So, the nature of the data structure and the nature of data search constrain the form within which data is collected and organised; the form of classification constrains the freedom to interrogate the data; the methods of interrogation constrain the ease in evaluating the results in terms of the desired goal. Equally, the nature of the interrogation or research process will influence the process of data file design. The design, interrogation and interpretation of the data file will now be discussed in more detail.

2.3.1 Data file design

Data bases can be employed to sort the small amounts of data that can be collected by the pupils themselves. The creation of a data file provides a good environment for the selection and organisation of the data collection activity, and involves the introduction of information handling skills such as deciding on field headings, classifying and coding, and entering the data into a data base. However, before going to the task of creating the data file, children need to have a purpose for classifying. As Underwood and Underwood (1990) have stated:

> *"The act of building a data file should require children to set goals or pose problems, collect and select data, and organise those data in such a way that the initial question can be solved" (Underwood and Underwood, 1990, page 77)*

In the same vein, Freeman and Levett (1985) say that it is the "hypothetico-deductive scientific model" which should be the framework which gives meaning to information handling, decisions about what data to collect, and how to define, structure and classify that data. All these activities should, they argue, have a bearing on a hypothesis.

This suggests there are intimate links between the information-seeking goals and the collection and organisation of data, and emphasises the importance for children to pose good research questions, because the definition of relevant data and the best ways for data organisation will depend on them.

There are good educational and psychological grounds for making the purpose, goal or information-seeking questions relate to the children's world. The initial organisation of the system determines the later usefulness of the file. The first step in classification is setting a goal for the classification. If this is not made clear to the students, the whole activity is likely to prove disappointing.

> *"Like the pen or the bicycle the computer data base should be seen as a tool with which to achieve a goal - it will not achieve the goal without a purpose of the activity being clearly specified and clearly related to concrete, understandable experiences" (Underwood and Underwood, 1990, page 82)*

Once students have decide what information they need for their projects, the data collection activity starts. Collection refers to the assembly of data from primary and secondary sources and its arrangement into a form appropriate for inclusion in the data file. Children have always had to collect and organise some kind of data for a paper or project. When using a data base, they have to decide the format and the content of the data file before they enter the data and use it. First of all, then, they need to understand the concepts of file, record and field, and then, to decide what information they should put into the data base; that is to select and organise the information that is relevant and sufficient to solve the problem, define the field headings, and finally enter the data.

One of the main problems in creating the file is the activity of coding data. Computers use coded data, that is reduced data, to facilitate the sorting procedures.There is evidence that the task of coding data entries is not always easy. Spavold (1989) carried out a study on 9 to 11 years old students' use of data bases, in which children were engaged in a local history project working with census material. She found that children, particularly the younger ones, faced difficulties in choosing the field names and reading the census enumerators's handwriting. She also found that the lack of keyboard skills was a major inhibitory factor in their work. Riggs (1990) points out that sometimes children feel frustrated when they discover the weakness of a badly planned data base, and suggests that since not all data bases allow the addition of categories or fields after the data file has been constructed, when establishing fields children could add three empty fields so that if the data file is set up and they realise that additional information should be placed in each record, one of the empty categories could be used.

The process of entering data into the computer is long and of doubtful educational value. Blease and Cohen (1990) report that children enjoyed it more when they had collected the information themselves, and that it was not as interesting as asking the computer questions. Hodson (1989) adds that boredom during the data entry stage is an ever present problem, especially with young children.

The creation of a data file is a rich data base activity which supports the development of organisation and classificatory skills. Success in building data files will be determined by the ability of the child to categorise and classify data, the response of the child to constraints regarding field and record size, the ability of the child to recognise the properties of data, and the fact that the file design affects the data file interrogation (Freeman & Levett, 1986).

2.3.2 Data file Interrogation

The interrogation of the file is another important data base activity carried out in the classroom. Hunter (1983) suggests a three-stage model for using data bases effectively across the curriculum, which entails three different levels of complexity, the interrogation of the file according to him occupying the first level. He distinguishes among: a) using data bases: students use a data base that has already been prepared to make enquiries and test hypothesis; b) building data bases: students create the data base by entering data through the computer keyboard into a format already designed by the teacher. They extract their data from reference sources, such as textbooks, encyclopaedias, etc.; c) designing data bases: students decide what information they need for their projects. Then they design the format and the content of the data themselves before they enter the data and use it. Underwood and Underwood (1990) report that primary school data bases are used mainly for the building of the data file, whereas in secondary schools and further education the emphasis is on information retrieval and hypothesis testing.

At the interrogation level, students are working with an already created data file to explore the information, search for particular records and fields or sort the data into order according to their own criteria and display it in different ways (graphs, tables, etc.). They can discover relationships and test hypotheses. The focus is now on posing the research questions to be asked and on specifying the query.

The quality of query formulation is the key to the effective questioning of a data base. The process of formulating the queries involves two fundamental aspects. The first has to do with the children's ability to ask worthwhile questions. The second lies in translating the question into the query language of the specific data base. Therefore, children need to be clear about what information they want to obtain from the search, and also about how to ask the computer so that it can understand.

Many authors have stressed the importance of children conducting their own investigations and asking their own questions, but less attention has been paid to the importance of children being aware of a problem that needs solving or of a question that requires an answer. However it seems that children do not often recognise the opportunities data bases offer and have trouble coming up with interesting questions to ask of the data base. Instruction organised around data bases should include direct attention to what makes a good question for the data base and how to think up such questions (Thomas, 1977; Perkins, 1985; Senior, 1989).

> *"Teaching a user a query language does not guarantee that they will use the data base in the most intelligent fashion to solve their problems. Additional training in their strategies for asking questions is worthwhile" (Thomas, 1977, page 180)*

Dillon et al (1989) add that when the students consider the data to be meaningful and capable of answering questions in which they are interested, motivation increases.

Once students have decided what they want to know from the data base, the query formulation process on the computer can start. This process involves transposing standard English questions into the command syntax of the machine using a variety of parameters to define the field values delimiting the search. One can construct any number of queries. To do this, those characteristics which must be fulfilled for particular fields have to be specified, and this is not always easy. Children do not always appreciate that the computer does not have the contextual clues that assist communication in normal verbal interaction. Queries must be correctly formulated and must explicitly contain all the features required to obtain the desired information.

At least five factors have been reported as important for the quality of queries. One is general computing skills (Beek et al, 1989; Spavold, 1989; Underwood, Spavold and Underwood, 1990). A second is an understanding of the field structure of the data base (Spavold, 1989; Underwood and Underwood, 1990; Shneiderman, 1978). A third is an understanding of the nature of data (Shneiderman, 1978). The fourth is an understanding of the use of Boolean logic in phrasing complex queries (Spavold, 1989; Langhorne et al, 1989; Underwood and Underwood, 1990; Bezanilla and Ogborn, 1992). The fifth, which is controversial, is the familiarity with the content of the data base (Hunter, 1983; Shneiderman, 1978; Underwood, Spavold and Underwood, 1990).

Some evidence of the importance of these five factors will now be reported.

2.3.2.1 Computing skills

In Underwood, Spavold and Underwood's (1990) study, computing skills, particularly basic keyboard knowledge were the dominant factor in the successful use of the data base. The authors report that if keyboard skills are not acquired rapidly, then loss of interest and motivation take place and learning is inhibited. Also in Spavold's (1989) study, lack of keyboard skills was a major inhibitory factor in children's work. These results highlight the need to introduce keyboard skills at a young age, if we want children to get the most of data base activities. A study by Beek et al (1989), however, showed that keyboard instruction produced little effect on the results.

2.3.2.2 Understanding of the data structure and search of the data base

To retrieve data effectively from the data base, the user must have some understanding of the nature of the data search and structure (see above, page 24, Durding et al, 1977; Broshey and Shneiderman, 1978; Underwood and Underwood, 1987, 1990; Ray, 1985; Marchionini and Shneiderman, 1990). Spavold (1989) also confirmed this fact. For the children in Spavold's study, understanding of the field structure was the most significant single factor in successful query formulation. Students who had not compiled the data base themselves had poorer understanding of field and record structures, and therefore, did not formulate queries so effectively or so quickly as those who had had the experience of the creation of the data base.

2.3.2.3 Understanding of the nature of data

In a paper by Shneiderman (1978) on the human factors of data base interactions, the author, when comparing natural and artificial language query facilities, found that although subjects posed approximately equal number of valid queries with either facility, natural language users made significantly more invalid queries which could not be answered from the data base. This indicates the fact that users know natural language syntax does not ensure that they know the semantics of data base interaction or the semantics of the information stored in the data base. Students need to understand the nature and types of data that can be contained and retrieved from a data base in order to formulate meaningful and answerable questions. Otherwise, children may have unrealistic expectations of the computer's power and may ask questions which involve value judgments and complex ideas which computers cannot and probably should not be relied upon to answer (Weizenbaum, 1976). Also, students might ask for some information which is not contained in a particular data base, thus wasting time and effort, while increasing frustration, or they might not be aware of the semantics of question asking and so not be able to think of what kind of questions could be answered by a data base. Shneiderman (1978) states one benefit of using a query language:

> *"By teaching users a concise and precise artificial language we are also teaching the semantics of question asking. Having a tool of a well-learned query language may enable users to compose complex queries which might not have occurred to them otherwise" (Shneiderman, 1978, page 428)*

2.3.2.4 Understanding of the use of Boolean logic in phrasing complex queries

It is often supposed that single-condition searches which require the use of options like "same as", "not same as", "greater than", etc. are easier than complex searches which require the use of the options AND and OR. In Spavold's study, the only pairs immediately comprehended and used by the students were "greater than/less than" and "equal to". Effective query writing other than one single-condition searches, needed constant help and checking. For the phrasing of complex queries, an understanding of the use of Boolean logic is required (Spavold, 1989; Langhorne et al, 1989; Underwood and Underwood, 1990; Bezanilla and Ogbom, 1992).

> *"Using a data base requires that students understand the concepts of keyword searching and Boolean connectors. It is in the skilful use of these two techniques that the true potency of a computerised data base is realised" (Langhorne et al, 1989, page 105)*

In one of the case studies reported by Underwood and Underwood (1990), children were unable to come to grips with the logic of constraining AND. Students felt that the more fields they specified the more records should be matched. Underwood and Underwood (1990) suggest this could be due to the children's lack of understanding of the concept of class inclusion/exclusion, or may well have been a problem of semantic associations.

The linguistic and the logical meanings of AND and OR are different. In ordinary speech "AND" can mean "AND IN ADDITION", adding an item, and not a constraining criterion, as is the case for the logical AND. A study by Reisner (1977) found that when people translated queries from English into a formal language, they did it phrase-by-phrase, leading to errors because of the AND/OR confusion.

Bezanilla and Ogborn (1992) (see Appendix 1) in a study of 13-14 year old students' performance on two parallel logical reasoning tasks related to data base searches, one requiring the interpretation of a logical sentence as a pattern of search, and the other requiring the construction of a logical sentence to achieve a desired pattern of search, found that children experience difficulty in interpreting and constructing logical sentences involving AND, OR and NOT, particularly negative OR sentences, tautologies and contradictions, and also that the constructing of a logical sentence is much more difficult than interpreting such a sentence. Spavold (1989) adds that children experience a high level of frustration when they try to convert from natural language to query language, finding it much easier to interpret queries and put them into English than to convert discursive inquiry into query syntax.

Bezanilla and Ogborn (1992) provide some reasons to suppose that children may have difficulties with the logical operators AND and OR (see Appendix 1). One is developmental, in the sense that the phrasing of complex logical enquiries requires a facility with propositional logic and an ability to consider different variables or conditions at a time, which according to some interpretations of Piaget is attained only in the stage of "formal operations". Therefore, it is to be expected that children's developmental stage will limit their performance on data base tasks. According to this, Langhorne et al (1989) suggest that data base skills would better be introduced at secondary school level.

> *"Teaching in depth data base skills to elementary children would be comparable to trying to teach the abstract notion of ratio in math to very young children. Therefore, junior high is an appropriate time for teaching students the core skill of using data bases" (Langhorne et al, 1989, page 89)*

Secondly, logical reasoning seems to depend on a subject's understanding of linguistic forms like AND, OR, IF, NOT, etc. and their different uses (Bower, 1977; McCawley, 1981; Evans, 1982; Braine and Rumain, 1983; Van Dijk, 1989). Children may have problems understanding the symmetry of the logical AND. In logic (p and q) is the same as (q and p), but in English, the order of clauses joined by AND is often not free.

There might also be a failure to distinguish AND as a logical operation from AND as a logical addition in the sense "AND AS WELL". For example, in a data base query, a sentence of the form "countries which are African and not African" should not be interpreted as "all the countries of the world" but as a contradiction. This suggests that it would be useful to study uses of AND and OR which lead to contradiction or tautology.

The operator OR can also present difficulties. It may be ambiguous between "inclusive" (either "A" or "B" or both) and "exclusive" senses (either "A" or "B" and not both), where in a data base, the meaning of OR is inclusive. Pragmatically, it seems that OR statements or queries are most often considered as being exclusive by listeners (Fillenbaum, 1977). Inclusive OR seems to be much more difficult for children to understand than exclusive OR (Braine, 1983).

Overall, this indicates that constructing logical data base queries is not a task to be taken for granted and that the teaching of data bases should include activities to explain the logical meaning of AND and OR. Special teaching on how to write queries should also be provided. The AND/OR confusion is only one way on which the English phrasing of a question may differ from the logical underlying structure.

To eliminate the confusion AND/OR operators from the language, researchers in the area of Human Computer Interaction have attempted to develop different query systems. One approach is to allow the user to state a question in natural English. A second approach is to require the user to state his question in a formal language system, but one that uses an English-like grammar and vocabulary (Reisner, Raymond and Chamberlin, 1975). A third approach is to require the user to state his question in a formal language system that does not attempt to appear "English-like", a nearly wordless system (Thomas and Gould, 1975). These systems are still at the level of experimentation, and obviously, have had no impact on education yet.

2.3.2.5 Familiarity with the content of the data base

Hunter (1983) reports that familiarity with the data base content also affects the query activity.

> *"The more familiar I am with my content, and with the format of my data base, the more efficient I can be in making my enquiries" (Hunter, 1983, page 324)*

Beek et al (1989) in a study on the effectiveness of a computer based training program for the retrieval of legal information assumed that legal knowledge was "prerequisite knowledge" for successful task performance. However, research by Underwood, Spavold and Underwood (1990) on undergraduates' use of relational data bases for research into local history shows that previous historical knowledge did not influence performance in a number of data base tasks.

Another kind of difficulty children may have when they begin to formulate queries is the division of their enquiry into those sections which would form the query and those which would form the output. In Spavold's (1989) study, children at first tended to ask for redundant output, to include fields forming query conditions in the output fields, but that once they had confidence in the system, this gradually ceased. The commands which summarise the query and the output are SEARCH and DISPLAY (QUERY and PRINT). The child is required to appreciate that a SEARCH is made of a field but that DISPLAYed fields can be other than those interrogated. Freeman and Levett (1986) state that the child's response to this appears to be hierarchical, in the sense that in early experience of data base use, the child makes a distinction, by preference, between DISPLAY and SEARCH. Brazier and Beishuizen (1988) add that the interface influences also the strategies employed by the students. In their study, simple DISPLAY was only used by the students using GRASS (menu-driven) and the use of a SEARCH to limit the number of items considered to the minimum was much more commonly used by QUEST (command-driven) students than by GRASS students.

2.3.3 Data file interpretation

> *"A data base is just a collection of data. No matter how usefully sorted and selected the information is, it needs 'humans' to analyse and evaluate it" (Watson, 1988, page 23)*

Once the results of an interrogation or search have been obtained, they constitute a criterion that can be used for judging and interpreting the data. Children can make inferences, reject or accept hypothesis, interpret results, etc. The results of the search may just be one step on the way to conclusion. Children should be taught not only to find information but also what to do with it once found (Langhorne et al, 1989).

As was mentioned previously, the most flexible data handling programs include the facility to display the data in a variety of graphic forms such as pie charts, scattergraphs or histograms. These can be very useful because they give children the chance to look at data in various ways and also encourage a meaningful interpretation of the data. Relationships between two fields can be found by correlating items graphically. Numeric descriptions of the variance and frequency of results can be obtained in many cases. Certain simple statistical procedures such as correlation or tests of differences may be applied. It requires skills and judgment to choose the appropriate methods for displaying results on the computer so that logical inferences may be drawn. Is a pie chart adequate, or would a cumulative frequency graph reveal more?

> *"Students need not only to be taught about the use of a data base but also to be given significant guidance about statistical practice and methodology if they are not to emerge with spurious, meaningless and often misleading results" (Welford, 1989, page 97)*

The ease with which the data can be transformed very rapidly in different methods of display means that different methods can be tried out to see if they are appropriate. The child has not only the capability of choosing, but also of testing and trialling. Seeing the data in different forms is said to make children eager to interpret the results more fully. It can give children the opportunity to become familiar with more techniques and therefore to make more rational judgments. A study by Welford (1989) which examines the use of small-scale data bases in introductory social science courses at degree level shows that those students who had used the data base facilities for generating graphics, data and trends, found themselves able to provide much more convincing evidence and justification than those who had simply reported the results of the raw statistics. He also reports that after using the data base, students did not become more confident with statistics, and suggests this could be due to the fact that data bases treat statistics in a "black-box" fashion, not allowing students to understand how results are produced.

Underwood (1985) reports that when children start testing an hypothesis with a data base, they often look for supportive rather than disconfirmatory evidence. The use of graphical representations may allow the exploration of data as a whole, and encourage hypothesis testing rather than hypothesis confirmation.

2.4 Data bases and the National Curriculum

The National Curriculum includes Technology as a foundation subject, and within it, it distinguishes between "design and technology capability" and "information technology capability". Programmes of study and statements of attainment regarding data bases use are made explicit under the "information technology capability" profile component. However, they are also contained in other subject areas of the curriculum.

> *"..... Information Technology Capability will be acquired through a range of subjects rather than through a single subject called Information Technology and it is likely to form an important part of all the National Curriculum subjects" (3.5 - Technology 5-16 in the National Curriculum, Nov. 1989, page 17)*

My immediate concern here is not so much to analyse the National Curriculum recommendations in depth but to examine in some detail what abilities regarding data base use are required for students, and to what extent they fit in with what has been said in the literature.

Data base attainment targets dictated by the National Curriculum cover two main aspects. First, there are those targets concerned with the skills involved in the use and operation of a data base. Second, there are those which emphasise children's understanding of personal, social and moral effects of data base use. These acquire more significance at the last levels, particularly at key stage 4.

Out of the 40 statements of attainment for information technology, five are strictly related to children's ability to use data bases. These are the following:
At level 3 of key stage 1 (5 to 7 years old) pupils should be able to *"collect data and enter it in a data base (whose structure may have been prepared in advance) and select and retrieve information from a data base"* ; at level 4 of key stage 2 (7 to 11), *"'amend and add to information in an existing database, to check its plausibility and interrogate it'"* ; at level 5, *"use a software package to create a computer database so that data can be captured, stored and retrieved"* , at level 7 of key stage 3 (11 to 14), *"select and interrogate a computer database to obtain information needed for a task"* ; and at level 8, *"select and use software to capture and store data, taking account of retrieval, ease of analysis and the types of presentation required" (Technology in the National Curriculum, NCC, 1990, pages 45-49).*

The design and interrogation of the data file are present at key stages 1, 2, and 3, however, importance has not been given to the interpretation of the data file. The data file design is first introduced at key stage 2 (7 to 11 years old), and the interrogation of the file at key stage 1. At this level children are expected to be able *"to retrieve information from a data base"* ; and the same is expected from 11 to 14 years old (Key stage 3), that is, *"to interrogate a computer data base"*. The problem is to know how distinct each statement is from the others, and how progression from level to level is to be achieved.

Particular subskills within the creation and interrogation of the file, such as the ability to organise and classify data, the coding activity, the query formulation process (simple and complex queries), the ability to deal with statistical tools, etc. have not been specified within the targets. The literature shows that children have difficulties in the phrasing of complex queries (there might be a developmental reason), and that factual retrieval is easier than the testing of ideas and relationships (Underwood and Underwood, 1990). Therefore, I would suggest that the attainment targets may not be specific enough to provide realistic goals and a clear basis for assessment.

2.5 Benefits claimed for the use of data bases

Several assumptions are implicit in the call for computer data base use. These assumptions sound logical and reasonable, but like so many aspects of computer use, little data exists beyond teacher anecdotes and classroom experiences to convincingly demonstrate that the assumptions hold true. In the previous sections of this chapter, some of these assumptions have been glimpsed. Now we will describe in some more detail the benefits of the use of data bases which are most generally claimed in the literature.

First, most of the researchers on this field agree that the major benefit of using data bases is that with them children have the opportunity to handle information more effectively. It is claimed that students are more efficient in completing data base activities if they use a computer than if they use traditional (paper and pencil) methods. The idea underneath this assumption is that data bases use the full potential of the machine itself. The fact that large quantities of data can be stored on a computer, the ready access and facility to interrogate such data, the possibility for information to be arranged in different useful ways, and the fact that data bases provide a structure is what makes data bases of great value (Freeman and Tagg, 1985; Degl' Innocenti and Ferraris, 1988; Hodson, 1988; Marran, 1986; Rawitsch, 1988; Underwood and Underwood, 1990; Jonassen, 1986; Hoelscher, 1986; Watson, 1988; Underwood, 1985; Knight and Timmins, 1986; Mittlefehldt, 1991).

To test this claim, some comparative studies of the use of computer versus printed material in solving some information handling tasks have been made. White (1987) compared the effects of structured learning activities with a computer data base, with the effects found with a relatively simple card filing system. The 665 secondary school students were divided into two groups, and both of them were confronted with the same curriculum specific data and a workbook to complete in ten days. The students' abilities to locate, gather, organise and evaluate information were measured in a test designed for this purpose, and the results showed that the students who were using computer data bases scored significantly higher on the tests than those using the card filing system. White ascribed this to the difference in structure encountered. In other words, he attributed the problem solving strategies employed for entering, organising and retrieving data to the structure imposed by the operation of the system. On the contrary, a study by Beek et al (1989) on the effectiveness of computer based versus traditional methods in solving legal information retrieval tasks shows that the computer based groups, although completing the searches quicker, did not retrieve relevant information better, nor did they demonstrate a higher efficiency or speed of using the command language. These results suggest that the practical usefulness of computer data bases remains to be demonstrated.

A second assumption is that data base analysis supports the development of higher level thinking skills required for effective problem solving throughout the curriculum (Rawitsch, 1988; Hodson, 1988; Degl' Innocenti and Ferraris, 1988; Jonassen, 1986; White, 1987; Watson and Strudler, 1989; Langhorne et al, 1989; Mendrinos, 1986; Olds, 1986; Smart, 1988). Through data base activities problem situations have to be analysed, questions formulated and searches planned. This means that students have to learn to identify problems that are significant, worth investigating and susceptible to systematic enquiry, and also that some questions or problems do not have a unique solution or right answer. Skills like classifying, comparing, hypothesizing, generalising and using Boolean logic are claimed to be enhanced through data base activities. It is claimed that data bases provide an ideal environment for "enquiry learning". Students can correct, format, rearrange and update information, show relationships, find trends and draw conclusions by asking questions to the computer.

> *"As students use a data base manager to store and organise information, they can generate and answer questions, formulate and test hypothesis, and critically evaluate the results of their enquiries" (Watson and Strudler, 1989, page 47)*

However, according to Watson and Strudler (1989) and Elder and White (1989), data bases only provide the medium for such enquiry. The rest depends on the well conceived implementation of sound teaching strategies that engage students in higher order thinking. This second assumption has its origin on Papert's (1970) view that computers, used as tools, might enhance thinking. Data bases are skill oriented and can stretch children's minds.

Beyond serving as a tool for teaching higher order thinking skills, it is claimed that the data base can also serve as an instructional vehicle for imparting a specific content. By interrogating and manipulating a data base, students can gain insight into the subject matter putting together relationships and displaying data in different ways (Langhorne et al, 1989; Hunter, 1983; Mendrinos, 1986; Rawitsch, 1988).

The fourth assumption is that data bases are a useful tool for providing motivation. Students are claimed to enjoy data base activities more if they use a computer (Mendrinos, 1986; Rawitsch, 1988; Underwood, 1986).

It is also suggested that learning activities through data bases reinforce and encourage group interaction and social participation (Hunter, 1987; Watson, 1986; Mittlefehldt, 1991; Croft, 1990), and they open up the classroom to more active learning, in the sense that they provide students with opportunities for asking questions in the classroom and influence the number and type of questions asked (Freeman and Tagg, 1985).

Another reason for introducing the use of data bases in schools concerns the needs of society. Data base use is an important carry-over skill for adult life, the development of which is best started during the school years (Langhorne et al, 1989; Rawitsch, 1988; Watson, 1988; McClelland, 1986; Sprowls, 1975; Hoelscher, 1986).

> *"If students learn to use a data base manager for storing and manipulating information, they will have taken a step forward preparing themselves for the information age, and they will have a tool to take with them" (Watson, 1988, page 21)*

2.6 A possible hierarchy of data base tasks

The previous sections have examined to what extent and in what ways the questioning of a data base is seen as an important cognitive goal. It has also been said that the complexity and type of data base tasks seem to depend upon the skills and cognitive development of the children. However, not all questions and tasks act as a stimulus to the higher cognitive skills of seeking relationships and reasoning about the world. The issues of the approach taken to questioning in the sense of whether the questions are seeking facts or testing ideas is of considerable importance in the assessment of the potential cognitive development of the children, and in the establishment of realistic and achievable goals regarding data base use.

Many authors agree that learning activities should be organised in sequential order to maintain continuity in learning (Inhelder and Piaget, 1958; Taba, 1971; Ausubel, 1968). This means that each learning activity should be built on previous learning, and each new activity should develop a basis for the performance required in the subsequent one. The National Curriculum also attempts to provide a framework for achieving continuity.

"Attainment targets and programmes of study allow for, and encourage progression whilst also accomodating differentiation. The 10 levels of attainment for each subject are designed to enable teachers, pupils and their parents to be clear about what is expected next of each pupil, and pupils to make identifiable progress through the levels at their own pace and from their own starting point" (NCC; From policy to practice, 1989, paragraph 4.14)

In developing thinking skills, this continuity seems to be especially important. They can not be learnt instantaneously, but take time and gradual building. As Taba (1971) states:

"A cumulative and a developmental learning sequence imply that learning activities proceed from the concrete to the abstract, from the experimentally close to the experimentally remote, and from tasks requiring simple thought processes to those requiring abstract and formal reasoning" (Taba, 1971, page 39)

The teaching of general information handling skills without using a computer seems to be arranged according to student's experience and development. Piaget and other developmentalists assert that mastery of basic operations, like seriating and classifying, is essential to cognition and prerequisite for success at later cognitive tasks (Inhelder and Piaget,1958). Beyer (1988) pointed out that the skills of classifying and seriating should be introduced in the first levels of education. He said that to those skills can be added instruction in comparing, contrasting and observing, and these should precede teaching in more complex forms of analysing, synthesising and evaluating.

Data bases, as compared with printed media, provide far more flexibility in manipulating data, so that a wide range of complexity of tasks is possible using the same data file. Tasks using a computer also need to be structured and sequenced. Hunter (1987) has suggested that children should start with activities which develop prerequisite skills, such as locating, classifying and interpreting information, before taking on tasks requiring more complex skills such as analysing, summarising, formulating hypotheses, synthesising, evaluating and communicating. Watson and

Strudler (1989) have made a distinction between learning tasks that involve the lower order skills delineated by Bloom (1956): knowledge (recalling information), comprehension (understanding literal messages) and application (using information in new situations), and those tasks which focus on the higher order thinking skills, such as analysis (breaking down the information to find relationships and connections), synthesis (putting together and organising the information to clarify the big picture), and evaluation (making personal judgments to explain the meaning and importance of the findings). Parker (1986) is a little more explicit when she states that the lower thinking skills with data base work might include entering data into a data base, retrieving factual information and using data bases to organise lists. The higher order thinking skills, on the other hand, might include determining what information is needed to test a hypothesis, reorganising and synthesising data to test ideas and find non-obvious relationships, discriminating between relevant and irrelevant information and drawing logical inferences and appropriate conclusions.

Riggs (1990) reports that in observing children working with a data base, it appears that they engage in different levels of activities, which seem to be hierarchically related. At level 1 we find *Browsing*: the child simply uses the data base to display individual records or lists from the data base. The information is displayed 'as is'. The child does not manipulate or select data to be displayed. At level 2, is the *Ordering* activity, in which the child changes the order of records in the data base according to a particular dimension, and then displays or prints them. Level 3 involves *Searching*: the child uses the data base to find records which have certain attributes or values. The last level is *Relating*: here the child looks for patterns and relations in the data base.

Underwood and Underwood (1990) have developed a classification of questions on the basis of the structure of the search procedure which was required to answer the query and on the purpose of the query. They distinguish among 1) Page data retrieval, in which the child simply asks the computer to reproduce on the screen specified blocks of the file with no manipulations. This mode is equivalent to the Browsing level defined by Riggs (1990); 2) Fact seeking retrieval, where the computer is used to manipulate data at two levels ("List all the known examples of..." and "Find the person which possesses the property of..."), and 3) Relational Retrieval where children look for relationships and patterns through the testing of hypothesis which can be either (i) implicit, implied by the type of questions asked, or (ii) explicit, clearly articulated hypothesis.

Phillips (1987) proposes working with a progressive data base consisting of five levels which are related to the skills of data handling. Level 1 simply offers the facility to view all the records. Level 2 allows one enquiry to be made, and the results of that enquiry are shown before further questions can be asked. Level 3 allows several enquiries to be made before the records are viewed. Up to five questions can be asked of the data at once, and the operators AND and OR can be selected (level 2 uses just the AND operator). Level 4 enables the selective graphic feature. Level 5 shows a great flexibility in the way enquiries are formulated.

Blease (1990) distinguishes between the formulation of queries involving one field, and those involving multiple fields.

Basically , what these studies suggest is that data base tasks, within the interrogation of the file activity, differ in complexity and demand. Easier tasks may require the retrieval of one or several specific items (factual information retrieval), and more complex tasks may require the seeking of relationships and the testing of ideas (relational retrieval).

In Brazier and Beishuizen (1989)'s study, instruction in the use of GRASS and QUEST consisted of five problems presented and solved step by step by the teacher.

Such problems addressed i) the retrieval of a characteristic of an object, ii) the retrieval of objects with a given characteristic, iii) the retrieval of objects with more than one given characteristic, iv) sorting objects on a characteristic, and v) finding a relation between two characteristics. The experimental task, however, consisted of 1) retrieval tasks, whose purpose was the exploration of the domain and the activation of the knowledge of the system, 2) reasoning tasks, which required the investigation of the relation between two or more attributes, and 3) tasks which required explorations of relation between all characteristics.

Taking into account Brazier and Beishuizen's steps on teaching data bases, and their distinction between retrieval tasks, reasoning tasks, and tasks which require exploration of relations between all fields, and the generally accepted view of the existence of factual retrieval tasks (requiring lower order thinking skills) and relational tasks (associated to higher order thinking skills), we can speculate on the existence of a hierarchy of data base tasks. It will be part of this research to examine the validity of such a hierarchy.

The hierarchy proposed includes two different levels of tasks depending on the reasoning skills they require:

1. Tasks which require the retrieval of specific data. This level has to do with tasks involved in the data file interrogation stage to which we referred previously. The interrogation of the file enables familiarisation with the nature of data, establishes the characteristics of the fields and their relationships to other fields, and encourages recognition and similarities between the records. It also favours the recognition and use of the different commands and options, the recognition of differences between SEARCH and DISPLAY, and between AND and OR conditions.

2. Tasks which require reasoning on the data retrieved. This level includes not only the interrogation but also the interpretation of the data file. These include investigation of the relations between two or more variables, test of hypotheses, explanation of the meaning of findings, understanding of different graphical representations, etc.

The two main tasks distinguished above can be subdivided into subtasks, giving six subtasks of different "hypothetical" levels of difficulty as shown below. The sequence of tasks is arranged in sequential steps, proceeding from the specific towards the more general, and from the simplest to the most complex.

<u>*A "logical" hierarchy of data base tasks*</u>

1. Tasks which require specification of simple retrieval statements

1.1. The retrieval of a given characteristic of a given object

1.2. The retrieval of objects with a given characteristic

1.3. The retrieval of objects with more than one given characteristic

1.4. Sorting objects numerically or alphabetically on a given characteristic

2. Tasks which require reasoning on the basis of data

2.1. Finding a relation between two characteristics

2.2. Exploration of relations between all the variables

The present research attempts to explore the possible existence of such a hierarchy in the children's work. The specific research questions will be stated and elaborated in Chapter 3. The hierarchy also constitutes the basis of the design of some of the instruments designed for the data collection. These will also be discussed in detail in the next chapter.

Chapter 3 RESEARCH QUESTIONS AND DESIGN OF INSTRUMENTS

3.1 The research questions

The present research focuses on two important elements of any teaching-learning process: the nature and structure of tasks and the skills required for them. Thus, it includes both an analysis of a number of data base tasks, and it involves questions about the skills and knowledge that students possess and which affect performance on such data base tasks.

The research has two main goals: first, to empirically test the "logical" hierarchy of data base tasks, theoretically defined in the previous chapter; and second to study those factors that the literature suggests are important for the quality of data base interrogation activity (see pages 32 to 38).

The study looks at two different age groups: children of 13 and 16, in order to see in what way the hypothesized hierarchy of data base tasks is apparent in the student's work, and if so, how it may depend on age. The role of the learner in relation to task difficulty is an important one. Therefore, attention is paid to those factors that the learner brings to the tasks. What is the students' understanding of the nature of data bases? Does it depend on age? And their understanding of logic? What other learner factors influence performance on data base tasks?.

The general question this research will try to answer could be stated then as follows:

> *Can we know what kind of different data base tasks children at a certain age and stage of development are able to achieve using a particular data base, and what prerequisites are needed for them?*

Having proposed a "logical" hierarchy of six different data base tasks (see page 50), the following research questions can be elaborated:

> *Can we know what kind of different data base tasks children of 13 and 16 years old are able to achieve using a particular data base (GRASS), and what prerequisites are needed for that?*
>
> *What are the difficulties of the various types of tasks? Are these different when children have to respond to questions already posed and when they decide what they want to know from the data base and form their own questions? What are the factors, related to the machine and related to the student, determining ease and difficulty?*
>
> *Does the "logical" structure of data base tasks produce a hierarchy which is apparent in the children's work? If so, does it depend on age?*

There are also a number of prior questions about children's understanding of data bases, and the skills and knowledge they possess:

> *What do children understand about data bases?*
>
> > *What is their knowledge relating to the general nature of data? What types of information do they think can be contained by and retrieved from data bases?*
> >
> > *What is their knowledge relating to the general nature of data base search and structure? How do they think information can be kept, organised and found in data bases?*
>
> *What is children's understanding of the logical operators AND, OR and NOT?*

The study includes some teaching on the use of a particular data base (GRASS); with pre- and post-tests. This work was designed to answer the following questions:

Does the teaching improve children's understanding of data bases?

Does the teaching improve their understanding of logic (logical connectors AND/OR)?

3.2 The instruments for data collection

The research questions formulated above evidently require the use of a varied mixture of different instruments and tasks for the investigation. These will now be described in detail.

3.2.1 To test the "logical" hierarchy of data base tasks

The materials used for testing the "logical" hierarchy of data base tasks consisted of:

- Two sets of data retrieval tasks
- A "report editing" task

3.2.1.1 Data retrieval tasks

Two different sets of exercises to be solved using a data base (GRASS) were designed and given to the children. Questions were identical in task demand, but not in content. Each set of exercises consisted of five questions of different levels of complexity according to the "logical" hierarchy of data base tasks. That is, there was one question on each of: the retrieval of a characteristic of an object (1.1), the retrieval of objects with a given characteristic (1.2), the retrieval of objects with more than one given characteristic (1.3), the sorting of objects numerically or alphabetically on a given characteristic (1.4), and the finding of a relation between two variables (2.1). The sixth subtask in the hierarchy: exploration of relations between all the variables (2.2), was not tested because of the excessive time required to attempt a solution. Children were not only asked to answer these questions about the content of the data file, but also to specify the way the search was made to get the answers.

The domain of the exercise was "Population of the World", which is addressed at different stages of education. The data file used, called POP, has information about 65 countries. POP holds 65 records with 18 fields in each record. Therefore, the whole data file contains 1170 different items of information. A description of the field structure of the data file is shown in Section A of Appendix 2.

The ten questions which constituted the two sets of exercises can be seen in Figure 3.1. Fig. 3.1 also reflects the level of each question within the theoretical hierarchy of tasks.

The two sets of exercises, in the same sequence as they were presented to the students, are shown in Section A of Appendix 3. It is important to notice that the sequence of tasks is different in each set of exercises, and that none of them follows the "logical" hierarchy of tasks theoretically defined.

LEVEL		EXERCISES 1	EXERCISES 2
1.1	Retrieval of a characteristic of an object	What is the population of Japan?	In which continent is Benin?
1.2	Retrieval of objects with one characteristic	In how many countries is the population smaller than 3 mill.?	How many countries belong to Europe?
1.3	Retrieval of objects with more than one characteristic	Which are the African countries whose growth of population is higher than 2.5 %?	How many countries have a birth rate under 15 per 1000 and a death rate under 11 per 1000?
1.4	Sorting data on one characteristic	Which is the country with the lowest death rate?	Which are the three most populated countries?
2.1	Finding a relationship between two variables	Do countries with a high GNP have a large percentage of people living in towns and cities?	Is there any relationship between birth and death rates?

Fig. 3.1 The two sets of exercises and corresponding levels in the "logical" hierarchy of tasks

The data base system used to interrogate the POP data file was GRASS (Graphics Searching and Sorting), a menu-driven program, developed by Newman College (1985). GRASS works on Nimbus and BBC machines, and it is very easy to operate because it displays instructions on the screen for moving around the menus and choosing what to do. A description of the eight different options included in the main interrogation menu is shown in Section B of Appendix 2.

Although some researchers (Brazier and Beishuizen,1988) found that command-driven programs like QUEST favour a more efficient data search than GRASS, in which the menu options, they say, seem to be difficult to remember and comprehend, GRASS was used in the experiment because it is one of the most currently available data bases in schools, and one of the very few educational data bases which allow graphical representations. Furthermore, it is very easy to learn and use. Together with POP data file, GRASS is included in an inservice pack, developed by MESU, to help teachers explore ways of using computers, and I was offered the opportunity to use them for the purpose of the research.

3.2.1.2 "Report editing" task

Requiring students to formulate queries in a constrained format, as the two sets of exercises do, does not really measure the student' s capacity to formulate queries to resolve problems. Thus, it was decided to offer students a problem where they had to formulate questions on their own.

To test not only the students' performance on answering given questions using a data base, but also their ability to form questions by themselves, the five different levels of data base tasks were also incorporated into a different task, which gave students a report to check and evaluate (see Section B of Appendix 3).

Unlike the Exercises, in which the questions themselves indicated the nature of the search, the Report did not give the student clues as to what to think about, what information to focus on, or how to search for it. It was intended to offer children the opportunity to find questions that could be asked about the text, to formulate them, and to find the answers. Thus, we can attempt to see if the hierarchy is reflected in responses, where the elements of the hierarchy are less evident in the task structure.

The report attempts to add an element of "real" investigation into the tasks. The task as presented can be seen below, in Figure 3.2. (The number in brackets corresponds to the position of each task in the "logical" hierarchy. The asterisk means that the particular statement is false). Obviously, these indicators were not present in the actual report given to the students.

In the Geography Department, we have found a Report titled "Population of the World", which contains particular information about different countries. It seems that some of the information contained in it is true, and some is false. We want you to check this report by using the data base and find out which statements are true and which are false.

Number each statement you check, and give it a √ (true) or X (false).
On the next page, explain how you checked each statement using the computer.

Report: "Population of the World"

The population of the world is very unevenly distributed. Some of the countries are sparsely populated, such as Nicaragua, New Zealand, Singapore and Jordan, with less than 4 million inhabitants (1.2). Some others are over populated, like India, which has more than 1000 million people (1.1*). Besides, the world population is not expanding at the same rate everywhere. In India, for instance, the population growth rate is high, 2.3 % per year (1.1), and this is made worse by the fact that its population is already the largest of the world (1.4*).

In some countries, the need to reduce the population growth rates has become urgent, and the reduction of birth rates seems to be the only means of doing it. The birth rate is over 50 in more than half of the African countries (1.3*), while in European countries, for instance, the birth rates are very low which may be because there is a high percentage of women using contraceptives (2.1).

The Gross National Product, that is, the average income of a country, gives an idea of its development. How well off countries are is reflected by the GNP. For example, USA, which is considered as the most developed country in the world, has the highest GNP per person (1.4), and on the contrary, the GNP in the African countries is not higher than 750$, if we exclude Mauritius, Tunisia and Zimbabwe (1.3).

Less developed areas, such as the three African countries Sudan, Uganda and Bangladesh (1.2*), have many interesting features. For example, in countries with low GNP, the infant mortality rates are lower than in rich countries (2.1*).

Fig. 3.2 The "report editing" task
(*) The number in brackets represents the level of that particular statement in the "logical" hierarchy of tasks. The asterisk means that particular statement is false.

The report contained information about Population of the World. It was designed from the POP data file. Students had to check the information contained in the report by using the data base. Ten statements were embedded in the text, five of which were true and five false. The statements were worded so that the five subtasks in the "logical" hierarchy were required to test them. The report task required students to identify what statements or pieces of information needed to be checked (English Attainment Target 2, level 5), and then to ask the appropriate question of the data base.

The ten statements embedded in the "report editing" task have been grouped according to their position in the "logical" hierarchy of tasks, and within it, in order of their appearance in the text, as shown below in Figure 3.3. The number in brackets again represents the level of the statement in the "logical" hierarchy of tasks.

The report was provided on paper and written answers were required. Students were asked to show the information they had selected to be checked, to specify the search path used to test the statements, and say whether the information checked was true or false.

LEVEL	STAT. No	STATEMENT
(1.1*)	2	Some others are over populated, like India, which has more than 1000 million people.
(1.1)	3	In India, for instance, the population growth rate is high, 2.3 % per year.
(1.2)	1	Some of the countries are sparsely populated, such as Nicaragua, New Zealand, Singapore and Jordan with less than 4 million inhabitants.
(1.2*)	9	Less developed areas, such as the three African countries Sudan, Uganda and Bangladesh, have many interesting features.
(1.3*)	5	The birth rate is over 50 in more than half of the African countries.
(1.3)	8	The GNP in the African countries is not higher than 750$, if we exclude Mauritius, Tunisia and Zimbabwe.
(1.4*)	4	Its population is already the largest of the world.
(1.4)	7	USA, which is considered as the most developed country in the world, has the highest GNP per person.
(2.1)	6	In European countries the birth rates are very low which may be because there is a high percentage of women using contraceptives.
(2.1*)	10	In countries with low GNP, the infant mortality rates are lower than in rich countries.

Fig. 3.3 Statements of the report arranged according to the "logical" hierarchy of tasks

The "report editing" task was thought of, not only as an appropriate research instrument, but also as a potentially useful diagnostic or teaching device for teachers to use. It represents the integration of data base tasks within a classroom type of task which could help students to develop the enabling skills they will need to interrogate a data base successfully. If it can be shown to work, then it may make some contribution to classroom practice. An advantage of the "report editing" task is that it can offer opportunities for work at all levels in the core curriculum subjects, and with classes of divergent abilities. Moreover, through this activity, students are making effective and sustained use of the information in the data base (English Attainment Target 2, levels 5 to 10) by selecting and retrieving data from the data base to identify the factual inconsistencies in the report. The report task can also become a diagnostic instrument for teachers who wish to know children's level of performance on data base tasks. In other words, it offers teachers the possibility to answer the following questions in relation to the students' data handling skills and knowledge: Can the student retrieve a characteristic of an object? (1.1), and sort data? (1.4), and find relationships between more than two variables? (2.2); Can he formulate questions to be answered by the data base?; What tasks do students find the most difficult?

Teachers can create their own reports and decide on those data base tasks that could or should be introduced or taught through the text/task. Children could also be encouraged to create their own reports from a particular data file. This could be done gradually. In the first stage, children could get familiar with the data file content, characteristics of fields and records, possible relations between fields, etc. In a second stage, students may be encouraged to write a paragraph with information about the file. They could include true and false statements. Thirdly, they could be made aware of the levels of the tasks they had introduced in their written passage, and of the possibilities of introducing different ones (the defined hierarchy of data base tasks could be made explicit), and finally, in stage 4, they could produce their own reports based on the data base skills and content they had acquired in stages one, two and three.

3.2.2 To test students' understanding of the logical operators AND and OR

In the previous chapter it was seen that the phrasing of complex data base queries requires children to understand and use the logical operators AND and OR, and that the difference between these operators is not always easily grasped by them. Results from the pilot study (see Chapter 4) also showed that task level 1.3 in the proposed hierarchy of tasks was the one that children found the most difficult, and it was thought this might be due to their lack of proper understanding of AND and OR as logical operators. The need to design some kind of instrument to measure children's understanding of such operators for the purpose of data base use was apparent. The problem, however, was how to do it. The approach taken in most studies of propositional reasoning, consisting of presenting AND and OR statements for children to translate, was not sufficient for the purpose of this study. The use of a data base requires not only the interpretation, but also the construction of AND / OR sentences. Therefore this "construction" element had to be present in the instrument. Another fundamental element of a data base is the structure of the data. The information retrieved from a data base is represented by a logical sentence, and also by a picture or structure. These two representations must be equivalent. It was decided that both representations (logical sentence and structure of data) should be part of the questionnaire as an attempt to simulate the way searches are done using a data base. After all, in a broad sense data base interrogation activity requires children both to interpret and build up logical sentences, and to organise and extract information from a particular structure of data.

Another aspect to be taken into consideration is that in most studies the logical rule is used to relate letters, numbers, shapes and colours (abstract material). We wanted to pose a more realistic problem which allowed the use of a logical operation which was appropriate in a particular "real life" situation.

In the end, two questionnaires were designed in order to test children's understanding and use of the logical operators AND and OR. Both consist of logical sentences, each linked to a grid, which pictures the structure of the data. The intention was to test the interpretation and construction of logical sentences of the logical forms

$p \,\&\, q \quad \sim p \,\&\, q \quad \sim p \,\&\, \sim q$

$p \vee q \quad \sim p \vee q \quad \sim p \vee \sim q$

and also the contradiction of the form (p & ~p), and the tautology (p v ~p), where 'p' and 'q' stand for propositions; '&' stands for logical conjunction (AND); 'v' stands for logical disjunction (OR); and '~' stands for negation (NOT).

The propositions were always chosen from sets of three alternatives. This was so as to clarify the role of negation. Suppose 'p' could be chosen from "is A", "is B" or "is C". If "is A" was chosen, then '~p' meant "is B" or "is C" or both. '~p' is therefore less specific than 'p', expressible only by one statement ("is A"). It should be noticed that if only two alternatives were used, then 'not p' could not be distinguished from the choice of the alternative.

For example, let us consider 'p' as "is B" and 'q' as "is 2". The 8 different logical relations between them mentioned above can be pictured as the following structures of data, represented on a grid (see Fig. 3.4):

p = "is B" q = "is 2"

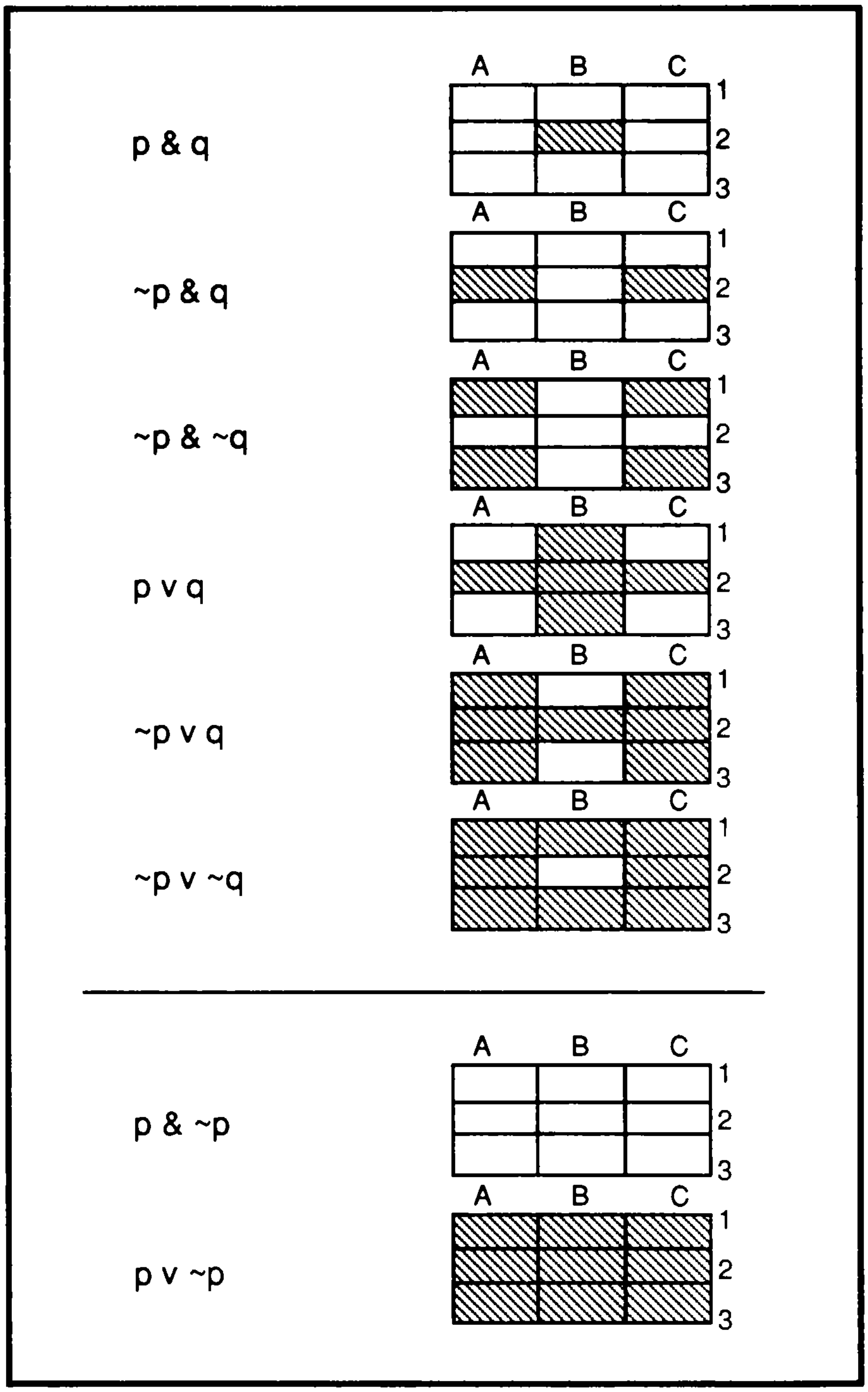


Fig. 3.4 Propositions and structures of data

The main difference between the two questionnaires is that in one of them, a logical sentence was given and students had to build up the right structure of the data in the grid *(Interpreting task)*. In the other, the structure of data was given, and students had to build the logical sentence which would allow them to get that picture or pattern of search *(Constructing task)*.

Both questionnaires were designed to be simple to use, to require little time to complete, and to be suitable for a wide age range. An advantage of these tasks is that they require non-verbal responses.

3.2.2.1 Interpreting task

The interpreting task contained eight different items or logical sentences, each of which involved one of the eight logical forms previously mentioned, including the contradiction and the tautology. 'P' referred to some electrical goods such as TV, CD players, and cameras. 'Q' referred to their country of origin (European, Japanese, American). Each sentence was linked to a grid, which represented shelves. Children were told that on the shelves there were leaflets with information about different electrical products, and they were asked to tick those shelves in which the particular goods specified in the sentence could be found. To make the instructions clearer, two examples with their solutions were presented.

An example of the two main components of the Interpreting task (logical sentence and structure) is shown below, in Fig. 3.5.:

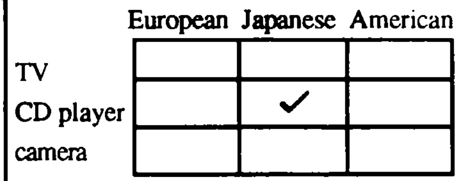
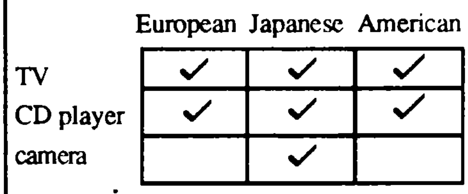
LOGICAL SENTENCE		PICTURE OR STRUCTURE
Linguistic form	Logical form	
Goods which are CD players AND which are Japanese	p & q	
Goods which are not cameras OR which are Japanese, or both	~p v q	

Fig. 3.5 Sentences and data structures for Interpreting task

3.2.2.2 Constructing task

The Constructing task consisted of eight different patterns of search, each of which represented the logical meaning of a given logical form. The logical sentences to achieve those patterns had to be constructed by the students. For this purpose, the components of the logical sentences (negation, 'p' and 'q', AND/OR) were presented in columns of alternatives and children were asked to circle the right alternative in each column, to build up the sentence which would allow them to obtain the information shown in the grid. In this case, the grid also represented shelves, this time holding passport information. 'P' referred to people (men, women and children), 'q' referred to their nationality (English, Scottish, Welsh). An example of how to construct the logical sentence corresponding to a data structure of the form (~p & q) was presented just after the instructions.

A description of the components of this questionnaire (Constructing Task) is shown in the following figure (Fig. 3.6):

<table>
<tr><th colspan="2">LOGICAL SENTENCE</th><th>PICTURE OR STRUCTURE</th></tr>
<tr><td>Linguistic form</td><td>Logical form</td><td rowspan="2">
<table>
<tr><td></td><td>English</td><td>Scottish</td><td>Welsh</td></tr>
<tr><td>Women</td><td>✓</td><td>✓</td><td></td></tr>
<tr><td>Men</td><td></td><td></td><td></td></tr>
<tr><td>Children</td><td></td><td></td><td></td></tr>
</table>
</td></tr>
<tr><td>All the people whose passport details are shown on the shelves :
<table>
<tr><td>ARE
ARE NOT</td><td>English
Scottish
Welsh
Women
Men
Children</td><td>AND
OR</td><td>ARE
ARE NOT</td><td>English
Scottish
Welsh
Women
Men
Children</td></tr>
</table>
</td><td>p & ~q</td></tr>
</table>

Fig. 3.6 Sentences and data structures for Constructing task

It would have been possible to present the nationalities in one list and the types of people in the other, instead of having both in each list. The reason for the presentation chosen was that in building a search sentence using a data base, all choices of fields to search are in fact available at every selection. Also, all are needed to construct tautologies and contradictions.

The order of presentation of the items was randomised in both questionnaires. Both the Interpreting and Constructing tasks can be seen in Section C of Appendix 3 .

3.2.3 To test students' understanding of the nature of data

To assess the student's understanding of the nature of data, a list of twenty questions was presented, and students were asked to say which questions could or could not be answered by retreiving data from a data base. Ten questions were intended to be answerable in this way and ten not. Those which were not, were related to constructs not amenable to data base analysis, such as Future, Context, Feelings, Judgment and Intention. Fig. 3.7 below shows these questions:

CONSTRUCT	QUESTION
Future	Will it rain more next year than it did this year? Is Margaret Thatcher / John Major going to be reelected in the next elections?
Context	Is it more convenient to go to Bristol by coach or by train? Do people wear smart clothes at a party?
Judgment	Which is the most objective British newspaper? Which was the best directed film last year?
Feelings	Do children enjoy using a computer? Do girls like getting flowers as a present?
Intention	Is the Soviet Union serious about arms reductions? Do football fans not really mean any harm?

Fig. 3.7 Questions not answerable by a data base

The ten questions answerable by a data base were grouped in three groups according to the actions they demand. These are retrieving data (one or several items), sorting data, and finding relationships between two variables. Out of the ten questions which can be answered by retrieving data from a data base, three belong to the category of simple retrieval, like *"Who is the Prime Minister of the British Government?"* (one fact retrieval), or *"How many cars were bought in England in 1988?"* (several objects retrieval). Another three belong to the category of sorting, like for example: *"Which is the highest mountain in North America?"*. The remaining four have to do with the finding of a relationship between two variables. Question 20 is an example: *"Is there a link between the number of inhabitants in a town and the number of schools?"* .

Children were asked to say which questions could or could not be answered by retrieving data from the data base. They just had to tick or put a cross against each question. The questionnaire can be seen in Section D of Appendix 3.

To some extent, the questionnaire was intended to see what expectations of the computer's power children with no experience of data base use had. The design of the questions, however, turned out not to be as simple as expected, and for example, questions like *"Do children enjoy using a computer?"*, may have led children to confusion. Computers do not have feelings, nor can they answer questions that involve feelings, but however, they can store information about peoples' feelings.

3.2.4 To test students' understanding of the nature of data search and structure

The literature shows that performance on data base tasks is affected by children's understanding of the way information is structured and found in a data base (see page 33). In order to operate a data base, students must have a mental model of it, however accurate or inaccurate. It was decided that an instrument to investigate children's ideas about the way data bases work had to be designed, and that such an instrument had to consist of open questions which would facilitate the elicitation of children's ideas.

Based on information about children's books, and using a card file as a metaphor, some questions were posed to find out how students imagined the computer could retrieve a characteristic of an object, objects with a given characteristic, and objects with more than one given characteristic. And also, how it could sort data and find relationships ("logical" hierarchy of data base tasks). The hierarchy of tasks was again implicitly present in this questionnaire. It was thought that in this way the data from this questionnaire and the data from the data base tasks (data retrieval tasks and "report editing" task) could be related in an attempt to get richer information on children's understanding and use of data bases. Introducing the hierarchy of tasks was also important to try to see whether children thought that all levels of tasks required a same search strategy, or whether the task characteristics, related and not related to the data base program (GRASS) determined the search path.

Fig. 3.8 shows the form of the questionnaire as well as the five questions involving the five levels in the hierarchy of data base tasks:

Suppose you have cards with information about children's books. Each card says:

- *Title of the book*
- *Kind of book (fairy tale, textbook, short-story, fable.....)*
- *Name of the author/s*
- *Date of publishing*
- *Publishing firm*
- *Number of pages*
- *Number of chapters/units*
- *Number of illustrations*
- *Type of binding*
- *Size*
- *Price*

The information on these cards is put into a data base on a computer.

We are going to tell the computer to find some information about the books and try to imagine how it might do it.

1. Suppose we tell the computer to find the name of the author of a book titled "Jane Eyre". How do you imagine it might try to do it? (1.1)

2. Suppose we tell the computer to find the titles of all the fairy tales. How do you imagine it might try to do it? (1.2)

3. Suppose we tell the computer to find all the textbooks which are cheaper than £5. What do you imagine it will do to try to find them? (1.3)

4. Suppose we ask the computer to put the books in order according to their price, ranging from the cheapest to the most expensive. How do you imagine it might try to do such a thing? (1.4)

5. Suppose we tell the computer to find out if the most expensive books usually have the most pages. How do you imagine it might try to do it? (2.1)

Fig. 3.8 Questions about children's understanding of the nature of data search and structure

In the last part of the questionnaire, children were asked to write down any other questions about the books, which they thought the computer could answer. The purpose of this last question was to measure student's ability to come up with interesting and meaningful questions to ask of the data base, and to find out what types of questions those are (factual - relational; simple - complex) and what levels in the hierarchy they belong to.

The actual questionnaire is shown in Section E of Appendix 3.

3.3 Overview of the research instruments

The research questions, stated in section 3.1, reflect the nature of the present study, with performance on data base tasks using a computer and prerequisite skills for good performance being the two main aspects explored. Since the nature of the information needed to address the research questions was varied, different instruments were used in the data collection.

Children's performance on data base tasks was measured through two different types of data base tasks. One consists of two sets of structured exercises posed in form of questions. Children needed to translate questions of different hypothetical levels of difficulty into queries. The other one is a non-structured data base task, the "report editing" task, which requires students to formulate questions, also of different difficulty levels, on their own. Data obtained from these tasks included responses to the questions or problems presented and specification of the strategies employed in interaction with the computer. In all, each child solved four questions/problems of each hierarchy level.

The rest of the information wanted was obtained through four specific questionnaires, which did not require computer work. Two questionnaires were designed to explore children's understanding of logical operators (the Interpreting and Constructing tasks). These required non-verbal responses and could be completed quickly. The other two were designed to test children's general understanding of data bases. One of these concerns the nature of data, and requires "yes" or "no" responses. The second concerns the nature of data search and structure. It consists of open-ended questions and requires a qualitative methodology.

The four questionnaires mentioned above were applied before and after having taught children how to use a data base (GRASS), and having given them the two types of data base tasks (exercises and "report editing" task) to be solved with the data base.

Chapter 4 PILOT STUDY, SAMPLE AND PROCEDURE

4.1 The pilot study

Some of the instruments of data collection described in chapter 3 were piloted with fifteen 16 year old students at a sixth form college in London. The aim of the pilot study was to get a feeling of how children reacted to the tasks, to see whether the tasks were clear enough to be comprehended by the students and whether they were actually testing what they were intended to.

4.1.1 How it was conducted

Students were involved in the experiment as a part of a course on the use of data bases in the Geography classroom. The school had fifteen Nimbus machines joined to a network, so children could work individually. All students had previously had some kind of experience of computer use, so they were familiarised with the use of the keyboard, but none of them had worked with any data base before. This information about the background and ability of the students was provided by the Geography teacher.

The pilot work was done during the months of May and June 1989, and it was carried out in three sessions, two of two hours and one one-hour sessions. The sessions were planned as shown in Table 4.1:

SESSION	DATE	ACTIVITIES
1 (2 hours)	17 / 5 / 89	• Questionnaires about children's general understanding of data bases: - about the nature of data - about the way information is kept, organised and found in a data base • Training
2 (1 hour)	22 / 5 / 89	• Two sets of data retrieval tasks - Exercises 1 - Exercises 2
3 (2 hour)	15 / 6 / 89	• "Report editing" task

Table 4.1 Sessions for data collection: Pilot study

During the first two hour session, students were given the two questionnaires related to their understanding about data bases. Before that, no instruction in what a data base is and how it works was given to them. Once the questionnaires were completed, the training began. Children moved to the computers and for an hour they were introduced to the basic features and characteristics of GRASS. After having given them some essential information about the system, like how to start the program, and the function of the arrow keys, Space Bar, Return and Escape, they were taught how to interrogate the data file with GRASS. Each option of the Menu was explained, and examples in the form of questions were presented and solved in interaction with the system. Such examples had not been planned before hand and I got the feeling that more preparation was needed. In order to keep children's attention, the training needed to be much more structured and systematic.

After the training, in the next session, students were given one of the two sets of exercises, and as they finished the first, the second was presented. Children worked individually, at their own pace. My observations suggested that at first they appeared to be more engrossed in exploring the different menu options than in answering the questions presented.

The next session (session 3) took place three weeks after the previous one. Since it was expected that in those three weeks children might have forgotten to some extent the system options and characteristics, time was provided to let them work on their own and become refamiliarised with GRASS. After that, the "report editing" task was administered and completed by eight students.

The two questionnaires on logic were not piloted. They were not even designed at that time. Also, as can be seen in Table 4.1, the questionnaires on children's understanding of data bases were not applied after the training.

The pilot proved to be very useful to determine the final length of the sessions and timetable. I realised that the training could be done in an hour, and that the two questionnaires on data bases could be completed in half an hour. I also learnt that the session in which children worked on the data retrieval tasks would be better if done just after the training, and that both should take place within the same session. It was also clear that the two sets of exercises as well as the "report editing" task could be completed in one hour.

In general I had expected the experimental work to take much longer. I was glad to see that GRASS could be grasped quite quickly and that children could complete data base tasks easily and quickly.

Fifteen students completed the two questionnaires on their understanding of data bases. Thirteen of them completed the two sets of data base retrieval tasks, but only eight were involved in the "report editing" task due to some difficulties in the school.

The questionnaires about logic (Interpreting and Constructing tasks) were informally piloted with six research students in March 1990, and the results showed that none did more than 4 items correctly in the Interpreting task, and that the task of constructing logical sentences was even more difficult.

4.1.2 Some preliminary results and modifications made to the instruments

The previous chapter reports the final instruments used in the research. The pilot study was of great use to decide on the final version of such instruments. In this section some preliminary results of the pilot study will be described, as well as the modifications made in the questionnaires and tasks on account of the results. As the number of students is very limited, statistical processing of the data has been avoided.

4.1.2.1 Questionnaire on childrens' understanding of the nature of data

No important alterations were made in this questionnaire. The results of the pilot study showed that all students responded to all questions, and most of them, with the exception of three, answered more than seventy five per cent of the questions correctly (of the fifteen students, twelve had a score above 15). All students completed the questionnaire in no more than fifteen minutes and they did not seem to have any problem in understanding the questions and attempting a response.

However, in relation to the types of information contained by data bases, the results showed some interesting features. Students scored on the average higher when recognising questions which can be answered by a data base (Average = 0.90) than in those which required recognising questions which cannot (Future, Judgment, Feelings and Context) (Average = 0.65). However, the average of correct answers to questions 5 and 13, which belong to the category of intention, was high (0.90) and after examining the questionnaire thoroughly, it was realised that they needed to be modified because they seemed to be too obvious. This could be due to the fact that they expressed an individual's intentions instead of joint or communal intentions. Questions 5 and 13 (*"Is Peter going to ask Mary to get married?"* and *"What did the teacher mean when he said 'get out'?"*) were converted into *"Is the Soviet Union serious about arms reductions?"* and *"Do football fans not really mean any harm?"* respectively.

Question 2, about Future (*"Will it rain more than usual next year?"*) and number 11, relating to Context (*"Is it easier to go to Bristol by coach or by train?"*) were slightly modified to make them clearer and to avoid confusion (see section D on Appendix 3).

Question 10 (*"Is Margaret Thatcher going to be reelected in the next elections?"*) had to be modified for the last group of students that took part in the experiment in the months of April and May of 1991 because at that time Margaret Thatcher had already resigned. In order to keep the maximum similarity between the old and the new questions, question 10 was simply changed into *"Is John Major going to be reelected in the next elections?"*.

4.1.2.2 Questionnaire on children's understanding of the nature of data search and structure

The results of this questionnaire were difficult to analyse. Although most students provided some kind of response to the questions, these were in many cases quite confusing, ambiguous and general, and therefore, difficult to categorise.

It seems that when having to explain how the computer could find some information from the data base, children had trouble coming up with new ideas or explanations, and their answers became in many cases, a repetition of the questions posed. It was thought that a reason for this could be that questions like *"Could the computer find* *the titles of all the textbooks? If yes - how; if no - why not"* were too rigid, not allowing children to imagine and visualise how the computer might work. Therefore, some modifications had to be made to avoid children feeling as if they were being tested and to let them express their own thinking more freely. The type of question shown above was transformed into a more flexible one, such as *"Suppose we tell the computer to find the titles of all the textbooks. How do you imagine it might try to do it?"*.

The content of the five questions involving the "logical" hierarchy of tasks remained the same, as well as the question which required children to write down any other questions about the data file that the computer could answer. Regarding this last question, it was interesting to see that half of the questions formed had to do with the finding of relationships between variables (level 2.1 in the hierarchy of tasks), and that no questions concerning the retrieval of objects with more than one given characteristic (level 1.3 in the hierarchy) were asked. The retrieval of a characteristic of an object (level 1.1) was also reflected in some of their responses.

4.1.2.3 Two sets of data retrieval tasks

No modifications were made in the two sets of data retrieval tasks. All the questions seemed to be well understood.

In order to analyse the results of the two sets of exercises, five different categories were established. Every question, regardless of its position in the "logical" hierarchy of tasks was analysed in the same way, according to the types of strategies used in solving the problem. The following criteria were used:

N = No answer
W = Wrong answer
A = Displaying all the information for all the countries

B = Displaying the specific characteristic or characteristics required for all the countries

C = Displaying just the information required

The basis for classification is illustrated by an example taken from one of the subtasks (level 1.1 in the "logical" hierarchy of tasks). The responses children gave to the question: *"What is the population of Japan?"* could be grouped under one of the following categories:

N - They may not answer the question

W - They may answer it in a wrong way. e.g.

Display
COUNTRY
Display

A - They may display all the information for all the countries

Display records
ALL FIELDS
Display

B - They may display the population of all the countries

Display
COUNTRY
POPULATION
Display

C - They may display just the population of Japan

Search data
COUNTRY
same as
(He will type in) JAPAN
display records
COUNTRY
POPULATION
Display

Two out of the fifteen students were excluded from this analysis because they answered none of the five questions of the second set of exercises, and no more than two of the first one.

The results of Exercises 1 showed that in general, strategy C was the most commonly used by the students. However, it seems that the use of a particular strategy to answer the exercises depends on the type of exercise and position of the question in the "logical" hierarchy of tasks. For example, the responses to question number 1 (level 1.1 - the retrieval of a characteristic of an object) show that ten out of thirteen students used strategy B (displaying the specific characteristic required for all the countries), and two used strategy C (displaying just the information required). On the contrary, nine of the total number of students used strategy C to answer question 3 (level 1.2 - the retrieval of objects with a given characteristic) and just one used strategy B. It is also important to notice that none used strategy A (displaying all the information for all the countries) at any of the five levels of tasks of the "logical" hierarchy.

Another interesting result is that the retrieval of objects with more than one given characteristic (level 1.3) seemed to be the most difficult task. This was apparent in the large number of no answers and wrong answers to this question. The complexity of this task could perhaps be explained by the need for children to understand and use properly the logical operators AND and OR. It is important to notice the parallelism between this result and that of the last question of the questionnaire about the way the information could be kept, organised and found in a data base. When children were asked to write questions about the books that the computer could answer, they also did not formulate any question concerning the retrieval of objects with more than one given characteristic (level 1.3 in the hierarchy). These results led to the design of the two questionnaires on children's understanding of the logical operators AND and OR.

The results of the second set of exercises seem to follow a similar pattern to those of the first set. Question 4 (level 1.3 in the hierarchy) seems to be also the hardest for children to answer. The least efficient strategy, strategy A, was only used by one student. Most of them employed strategies B and C. The number of no answers decreased substantially. Although the answers given by the students did not vary to a large extent, it can be said that in general, students scored slightly better in the second set of exercises than in the first one. After having grouped strategies B and C under the category of "good response", and no answers (N), wrong ones (W) and strategy A under the category of "poor response", this slight difference in performance can be better appreciated.

Table 4.2 shows a description of the results of Exercises 1 and Exercises 2.

	Task level	Quest. No.	ΣN	ΣW	ΣA	ΣB	ΣC	Poor Resp.	Good Resp.
EXERCISES 1	1.1	1	0	1	0	10	2	1	12
	1.2	3	3	1	0	1	9	3	10
	1.3	5	1	3	0	5	2	6	7
	1.4	2	2	2	0	2	8	3	10
	2.1	4	1	2	0	3	6	4	9
							Σ	17	48
EXERCISES 2	Task level	Quest. No.	ΣN	ΣW	ΣA	ΣB	ΣC	Poor Resp.	Good Resp.
	1.1	2	0	0	1	6	6	1	12
	1.2	5	0	0	1	6	6	1	12
	1.3	4	1	5	0	4	3	6	7
	1.4	1	0	1	0	3	9	1	12
	2.1	3	0	1	0	4	8	1	12
							Σ	11	54

Table 4.2 Results of the two sets of data retrieval tasks: Pilot study

4.1.2.4 Detecting and correcting errors in a Report

Only 8 out of 15 students responded to the report task, and of them two were excluded from the analysis because they either did not check most of the statements or did not explain the search path they used to reach the answers. The analysis of the report consisted of two parts:

- Number and type of statements children checked

- Ways of checking or strategies used for testing the statements

Concerning the children's ability to recognise the information that needed to be checked, four students detected and tested the total number of statements (ten) included in the report, and just two students tested less than fifty per cent of the total information. The activity of finding the statements they had to test seemed to be independent of the position these statements had in the "logical" hierarchy of tasks. For all the statements, the number of students detecting the statements correctly varied between 5 and 6 out of 6.

The same criteria used for the analysis of the data retrieval tasks were employed to analyse the results of the report task.

Most students employed strategies B and C and none used strategy A in any task. The proportion of wrong answers was smaller as compared with the exercises, and in general, it seemed that the level of performance was slightly higher in this task than in the sets of exercises. One possible explanation for this fact could be improved memory of the system by the students due to reinforcement by repeated use.

From the results of the pilot study, it was realised that the report task needed to be improved. The last two statements (*"Medical facilities are generally poorer in the less developed areas. In countries with low GNP, the infant mortality rates are lower than in rich countries (2.1*). This is the case of the three African countries: Sudan, Uganda and Bangladesh" (1.2*)*) were, as above too close together to be seen as distinct. Statement 10 (*"This is the case in three African countries: Sudan, Uganda and Bangladesh"*) was not interpreted separately as the testing of whether a number of countries (Sudan, Uganda and Bangladesh) share one same characteristic (being African) (level 1.2 in the hierarchy - the retrieval of objects with a given characteristic), but in connexion with the previous statement, statement 9.

Statement 8 (*"The GNP in the African countries is not higher than $750, if we exclude Mauritius and Tunisia"*) had to be changed too because the information given was not that which was desired. An item of information was lacking (Zimbabwe, whose GNP = 760 $/per.), and statement 8, which was supposed to be true, became false.

The first statement (*"Some of the countries are sparsely populated, such as Nicaragua and New Zealand, with just 3.2 millions of inhabitants"*) was also modified and expressed in a different way to make it more significant. This statement corresponds to the retrieval of objects with a given characteristic (level 1.2). It seemed likely that the more objects which were required to be retrieved, the less students might interpret the statement as the retrieval of a characteristic of an object (level 1.1) to be done twice. Therefore, it was decided to include four objects (countries) rather than just two in the modified version of the report. The final "report editing" task, after having made the modifications, is shown in Section B of Appendix 3.

4.2 The main study

4.2.1 Description of the sample

Since one aim of the research was to see whether a theoretically defined hierarchy of data base tasks was apparent as such in the students' work, and if so how it depended on age; and also, since this study of children's use of data bases tried to see to what extent children's cognitive development, in terms of reasoning skills related to using data bases (understanding of Boolean logic) affected performance on such data base tasks, the population tested needed to include subjects of different age groups.

The research looked at two different age groups: twenty five children of 13 years of age, taken from a secondary school in Hertfordshire (School A), and twenty five 16 year olds, taken from two different schools in London (Schools B and C). Schools B and C were very similar in nature, both urban and with mixed ability children. School A came from a rural area and according to the school had predominantly average to above average ability pupils.

The selection of the age groups was not done arbitrarily. At first, it was planned to work with children of at least three ages in order to obtain a broader picture of children's use of data bases (GRASS in particular) and of the factors that influence the interrogation activity, in an attempt to see whether performance bench-marks for certain data base tasks could be defined. It was also hoped to incorporate a group of younger children (10-11 years old) to see to what extent children's developmental stages, as suggested by Piaget, would limit their performance on data base tasks, particularly in those which required the use of the logical operators AND and OR. However, access to schools in which to conduct the research turned to be very difficult for the reasons explained below, and in the end just two age groups could be examined.

The selection of schools was not easy and was done according to certain criteria. First, since students were required to work individually, schools needed to have as many computers as students within each group, that is a minimum of 10 to 15 computers (either BBC or Nimbus). Most primary schools do not have such good computer facilities.

Second, schools had to agree to provide pupils for a minimum of three one-hour sessions and one two-hour session. Obviously some schools did not allow their students to take part in the experiment because that meant students had to miss five hours of their current learning of the particular curriculum subject, which could interfere with their performance in examinations. Teachers and Heads of Department had to be convinced that their students would find the work interesting and that what they learnt could be applied to their own work. This became a major difficulty in the school selection process.

Third, since students would be tested in their general understanding about data bases before and after being taught about data bases and trained in the use of a particular data base (GRASS), it was important that students had not worked with any data base before. At the same time, it was required that they had some familiarity with the use of the computer keyboard.

Fourth, since the research involved the investigation of a number of aspects regarding data base use, and the application of several tasks and questionnaires; and because students were pre-tested and post-tested, it was important to make quite sure that all students would attend all experimental sessions if the research was going to be valid. However, a number of students missed one or more sessions, and they were not included in the analysis. This interfered with the sample size.

Four groups and two age groups were studied. These were 13 year olds (Groups 1 and 2), and 16 (Groups 3 and 4). The table below describes schools and pupils participating in the study.

Group	School	Age (years)	No.of Boys	No.of Girls	Total No. students
1	A	13	6	8	14
2	A	13	6	5	11
3	B	16	5	8	13
4	C	16	9	3	12

Table 4.3: Description of the sample

The twenty five pupils in School A were divided into two groups of thirteen and twelve for the work.

Since schools B and C were very similar in nature, for the purpose of the analysis students were grouped simply according to age.

4.2.2 Procedure and administrative details

The nature of the work to be carried out was discussed before hand with each school. In schools A and C the work was done within the Information Technology classroom where the teaching of data bases, and GRASS in particular could be included easily into the teacher's teaching scheme. The school A teacher had never worked with GRASS before, but he was very willing to learn about the possibilities and learning strategies GRASS could offer. School C teacher was quite familiar with GRASS but had never introduced it to his students. To teach the use of data bases to the students of previous years, he had always used QUEST. He thought this could be a good opportunity to learn more about GRASS and to see how children reacted to it compared to QUEST.

In school B, the experimental work was done within the Geography classroom. The Geography teacher was very keen to take part in the research work, as she saw it as an opportunity of starting to introduce the use of computers, and of GRASS in particular in Geography. She was particularly interested in children's attitudes and responses to the computer graphs.

The same experimental work was done at each school. It was hoped that the data collected would give some indication of the similarities and differences between the age groups. The study was carried out in four sessions, three of one hour and one two-hour session. During these sessions, students were:

(i) tested on their understanding of the logical connectors AND and OR, and on their general knowledge about data bases

(ii) trained in the use of GRASS

(iii) tested on the use of GRASS in solving data bases tasks of hypothetical different level of difficulty (hierarchy) expressed in form of questions (Exercises), and on their ability to form questions to be answered by the data base ("report editing" task)

(iv) Tested again on their understanding of Boolean logic and on their general knowledge of data bases

Table 4.4 shows a schematic overview of the experimental procedure

SESSION	ACTIVITIES
1 (one hour)	• Logic questionnaires: . Interpreting Task . Constructing Task • Questionnaires on children's understanding of data bases: . about Nature of Data (types of information that can be contained by data bases) . about Nature of Data Search and Structure (ways information can be kept, organised and found in a data base)
2 (two hours)	• Training: . general introduction to GRASS and explanation of the system (1 hour) • Exercises 1 (30 minutes) • Exercises 2 (30 minutes)
3 (one hour)	• "report editing" task
4 (one hour)	• Logic questionnaires: . Interpreting Task . Constructing Task • Questionnaires on children's general understanding of data bases: . Nature of Data . Nature of Data Search and Structure

Table 4.4 Description of the experimental procedure

A more detailed description of the work conducted in each of the four sessions is given below.

<u>Session 1</u>

During the first one hour session, students were given the two questionnaires about logic, and the two questionnaires related to their understanding about data bases, one at a time. As soon as they completed the first, the second one was given. Children worked individually, and the researcher tried to make sure they did not talk amongst themselves. In order to give them confidence and to favour the free expression of their own ideas, they were told that there were no right or wrong responses, and that they would not be assessed. Children seemed not to find it difficult to respond to the questionnaires, with perhaps the questionnaire on their understanding of the nature of data search and structure being the one which gave them most trouble. Even so, all students completed the four questionnaires in one hour. Before that, no instruction in what a data base is and how it works was given to them.

Session 2

In session 2, the training took place. In the first place, children were taught what a data base is, and introduced to the notions of record and field. This was done very briefly with the aid of index cards. Then, the screen layout was explained, as well as the use of the arrow keys, Space Bar, Return and Escape. A description of the function of these keys was written on the whiteboard and was left there until the session finished. Next, children were taught how to interrogate the POP data file with GRASS. The menu options DISPLAY, SEARCH and SORT were explained. and examples in form of questions were presented. While I was explaining how to reach the answers, I was writing the search paths on the whiteboard. Children were following my instructions interacting with their computers. Less attention was paid to the option GRAPHS and FIGURES, since at the end of the training session, children were just exploring the menu options on their own, and not following my instructions so strictly.

The aim of the training was to acquaint students with the main commands of GRASS. As said before, all the students had previously had some kind of experience of computer use, so they were familiar with the use of the keyboard, but none of them had previously worked with any data base.

I myself developed and gave the training to the four groups. The training aimed to be exactly the same for all groups. The same content, and the same examples given in the same order were introduced to all students. I also tried to use the same kind of vocabulary for all groups. A full and detailed description of the training can be found in Appendix 4.

During the training I could observe that some children were not strictly following my instructions, but instead, were freely exploring the program by themselves. In general, their attitudes towards the computer and towards GRASS in particular were very positive. They did not show any kind of initial fear or bemusement.

Generally, children grasped the meaning of DISPLAY and SORT very quickly. SEARCH was more difficult to understand, particularly when it was used for the construction of complex searches. The distinction between OR and AND had to be emphasised and explained very carefully more than once. Even so, once the training had finished, some students still could not understand the difference between the two logical operators. They also tended to forget that every time they wanted to do a new search, they had to select the option "start a new search". GRAPHS AND FIGURES was the last menu option introduced to the students and it was not taught in great depth. Although children seemed to be happy and confident with this option, maybe some more time would have been beneficial.

After the training, students were given one of the two sets of exercises, and as they finished the first, the second was presented. The paper with the description of the POP data file structure (see Appendix 2) was also given to the students to help them in solving the tasks. I supervised the eight computer sessions.

A limit of time to answer the exercises was not imposed, but all students completed the two sets of exercises in one hour approximately. They rarely asked for help, although sometimes they asked for reassurance when they asked me: "Is this all right?"

Overall, children did not show too much enthusiasm in working with GRASS. However, a few seemed to enjoy their session a lot to the extent that, after having finished their work, they remained in the classroom playing with GRASS.

Session 3

Session 3 took place one week after the previous one, except for Group 1, for which it took place three weeks after (see Table 4.5 below for a description of the research timetable). Since it was expected that in those three weeks children might have forgotten to some extent the options and characteristics of the system, time was provided to let them work on their own and become refamiliarised with GRASS. After that, the report task was administered and completed by all students in no more than one hour .

Pupils seemed to enjoy this particular session the most. This could be due to the nature of the "report editing" task itself, which is more intellectually demanding, and more "real". Through it, children may have recognised the usefulness of data bases better. Another reason could be that children were more familiar with the system commands, and this acted as a motivating device.

Session	Group 1	Group 2	Group 3	Group 4
1	11 / 5 / 90	11 / 5 / 90	2 / 7 / 90	19 / 4 / 91
2	18 / 5 / 90	15 / 6 / 90	3 / 7 / 90	26 / 4 / 91
3	8 / 6 / 90	22 / 6 / 90	6 / 7 / 90	3 / 5 / 91
4	15 / 6 / 90	29 / 6 / 90	9 / 7 / 90	10 / 5 / 91

Table 4.5 Timetable of the experimental work

Session 4

This session was similar to Session 1. The same questionnaires were given again in the same order. The aim was to see whether children's general understanding of data bases and their understanding of the logical operators AND and OR had improved after having given them some training in searching information with GRASS, and the opportunity to solve some data base tasks of different levels of difficulty. It was also hoped to see whether learning to use a particular data base requires a previous understanding of the nature of data and of the way data is organised and found in the data base, or whether one can become a skilled data base operator with little or no conceptual understanding of the program or of the processes he is involved in.

I myself directed all sessions for all groups with the exception of Session 4 for Group 2, which was run by the I.T. teacher.

Chapter 5 CHILDREN'S PERFORMANCE ON SOME DATA BASE TASKS

5.1 A First Level of Analysis: Responses and Strategies

Two aspects were considered in the analysis of the data retrieval and "report editing" tasks. First, the correctness of the responses made and of the strategies used, and second, the nature of the strategies used in solving the tasks. Both product and process were included, then, in the analysis. (The tasks can be found in sections A and B of Appendix 3).

Answers were categorised as: Right, Wrong, and No response, both for the answers to the questions and for the specification of the ways children got those answers.

Three categories were used to analyse the strategies used by the children in the searching process:

- Strategy A: Displaying all the information for all the records
- Strategy B: Displaying the specific characteristics or fields required for all the records
- Strategy C: Displaying just the information required, after having done a search.

The use of strategy C means that the data base is being used to its full advantage, and is the most efficient. Strategy B means that the computer displays particular fields and children have to find themselves the information required by checking, counting, matching and/or comparing the information displayed. In other words, the computer does part of the job but some work is left for the child. The least efficient strategy, strategy A, means that the computer displays all the information contained in the data base as it appears in books and written documents, and there is a lot of additional work the child has to do to find the answers. Strategy B (Display) requires the child only to specify the fields in which the result is to be found. However, strategy C (Search) requires specifying possible values of data in those fields. In other words, a "Search" operation needs the child to think of a relationship in which values for the Search will be specified. A "Display" operation presents possible values for consideration. Marchionini and Shneiderman (1990) refer to strategies B and C as "scan and select" strategy and "analytical search" strategy respectively.

An example of how the responses to one of the tasks (level 1.1) were categorised was provided in Chapter 4 (page 78). Doing a partial search to answer those questions which required the retrieval of objects with more than one given characteristic (level 1.3 in the hierarchy) was also categorised under Strategy C. For example, a complete search for the question *"Which are the African countries whose growth of population is higher than 2.5%"* would be:

Search data
POPGROWTH
is greater than
(He will type in) 2.5
Narrow down the search (AND)
CONTINENT
same as
(He will type in) Africa
Display these records
COUNTRY
CONTINENT
POPGROWTH
Display

It involves two searched fields joined by the operator AND. A partial search could be, however, either:

Search data
CONTINENT
same as
(He will type in) Africa
Display these records
COUNTRY
CONTINENT

POPGROWTH
Display

or

Search data
POPGROWTH
greater than
(He will type in) 2.5
Display these records
COUNTRY
POPGROWTH
CONTINENT
Display

where a search is done on one field and AND is not used to join the fields. It was important to see at this point whether children appreciated that a search is made of a field but that displayed fields can be other than those interrogated.

As will be seen below complete searches for this particular type of tasks are not common. It seems that children have difficulties to use the logical operator AND to answer tasks of level 1.3.

For task level 2.1 (the testing of relationships), strategy B consists of the display of the two variables for all the records, and strategy C usually includes a scattergraph representation of the two variables as seen in the example below (Question 4: *"Do countries with a high GNP have a large percentage of people living in town and cities?"*):

Graphs and figures
GNP/PERSON
Scattergraph
URBANPOP

Less commonly, strategy C includes splitting the data file up by searching for say all countries whose GNP is less than $5000, create graphs and find the urban population for these countries, and then search for countries whose GNP is greater than $5000 and get more information.

It is hoped that the examination of the strategies used by the students will allow me to make some inferences about their understanding of the types of operations data bases can do and the ways they do it, and also to see whether differences in effectiveness (success in finding particular information) are a result of the types of strategies used. It is also hoped to glimpse whether the use of one or another strategy is subject to individual differences or to task characteristics, or to both.

5.2 Analysis of the results of the two sets of data retrieval tasks

For ease of discussion, the younger pupils will be discussed first, followed by the older pupils. Analysis of the correctness of the responses and of the strategies used, as well as of the common type of errors children made will be presented first, followed by examination of the nature of the strategies used to solve the problems.

5.2.1 Younger students

In relation to the first set of exercises, the results in Table 5.1 show that almost all students answered correctly the questions concerning the retrieval of a characteristic of an object, the retrieval of objects with a given characteristic (levels 1.1 and 1.2 respectively in the hypothesized hierarchy), and also the question which required sorting data (level 1.4 in the hierarchy). However, some children, although they gave a correct answer to the question, failed in the description of the path or strategy used to solve it. The most common type of error was not displaying the information once the search had been done. It should be noticed, however, that for the questions asking 'How many', the data does not need to be displayed. After a search has been done, the computer shows on the screen the number of records that match that question in comparison to the total number of records in the data file. This is not enough for the 'Which' or 'What' questions, which require the information to be displayed.

EXERCISES 1 - Responses -						
TASK	RESPONSE			STRATEGY		
LEVEL	R	W	N	R	W	N
1.1	24	0	1	22	1	2
1.2	22	3	0	23	1	1
1.3	14	5	6	6	12	7
1.4	25	0	0	22	3	0
2.1	14	4	7	13	8	4
Σ	99	12	14	86	25	14

R = Right
W = Wrong
N = No Response

Table 5.1 Responses to the first set of data retrieval tasks - Exercises 1: 13 year olds

Just about half the students, 14 out of 25, answered correctly the question involving the retrieval of objects with more than one given characteristic (level 1.3 in the hierarchy), but of those, only 6 specified a correct strategy. Some of the rest either focused just on one of the conditions or did not use the menu options AND/OR to connect the two characteristics. Some others chose the wrong field to search (for example, POPULATION, instead of POPGROWTH) or did not display the information after the search had been pursued. One student used the logical operator OR instead of AND, adding to the search instead of narrowing it down. It should be noticed that all students who answered the question wrongly had used Strategy C, and that of those who had used Strategy B, only 2 got a correct answer (see Table 5.2). Slightly more than 20% of the students did not give any kind of response to this question. Six students answered this question correctly without describing any strategy. This could be explained by the possible difficulty children may have in verbalising those data base operations which involve the use of logic. They can perform them, but then they are unable to describe them.

		Strategies						Strategies			
		A	B	C	N			A	B	C	N
	R		2	6	6		R		2	4	
Response	W			5		Strategy	W		2	10	
	N		3	2	1		N				7

Table 5.2 Relationship between responses and nature of strategies used for task 1.3 (retrieval of objects with more than one characteristic) - Exercises 1: 13 year olds

About half of the students answered correctly the question involving the finding of a relation between two variables (level 2.1 in the hierarchy). The most common type of error was the wrong selection of the fields to search. To answer question 4: *"Do countries with a high GNP have a large percentage of people living in towns and cities?"*, children often displayed the fields POPULATION and GNP, instead of GNP and URBANPOP. Some others focused on only one variable, and displayed just GNP. The number of no responses to this question is also quite high. Even after having pursued a correct strategy, it seems that some children had difficulties in saying whether there was or was not a relationship, or whether it was positive or negative.

In relation to the strategies used by the students (Table 5.3), no one used strategy A (displaying all the information for all the countries) at any of the five levels of tasks. Strategies B and C are the two most frequent, C being a little more common than B. However, it seems that the use of one or another strategy depends to some extent on the type of task and on the position of the question in the "logical" hierarchy. For example, regarding Question 1: *"What is the population of Japan?"* (level 1.1 in the hierarchy), 16 out of 25 students used strategy B (displaying the specific characteristic required for all the countries), and 7 used strategy C (displaying just the information required). By contrast, 18 used strategy C to answer Question 3: *"In how many countries is the population smaller than 3 million?"* (Level 1.2. The retrieval of objects with a given characteristic) and just 5 used strategy B. It seems that children understand that finding an item from a list is a lot easier than finding several, and that for the latter, doing a search (not just displaying and counting) is more efficient.

EXERCISES 1 -- Strategies -				
TASK	STRATEGIES			
LEVEL	A	B	C	N
1.1	0	16	7	2
1.2	0	5	18	2
1.3	0	4	14	7
1.4	0	11	14	0
2.1	0	11	10	4
Σ	0	47	63	15

Table 5.3 Nature of strategies used: Exercises 1: 13 year olds

Strategy C was also the most commonly used, although without success, to answer question number 5 (*"Which are the African countries whose growth of population is higher than 2.5%?"*), for which the use of the logical operator AND is required (4 students used strategy B and 14 used strategy C). As was said previously, many students did not use Strategy C properly (see Table 5.2). Attention paid to just one condition and searching for the two fields separately (CONTINENT or POPGROWTH) were the most common type of errors. Only 4 students did a correct complete search, in which the operator AND is used to join the two conditions:*"being African* and *having a population growth higher than 2.5 %.".*

Strategies B and C were used almost without discrimination to solve tasks of levels 1.4 and 2.1 (sorting data and finding relationships respectively).

Looking at individual's behaviours, just two students used strategy B and another two strategy C for the five questions (another four students used strategy C consistently but they did not answer one or more questions).

The results of the second set of exercises, done shortly after the first, seem to follow a quite similar pattern to those of the first set (see Table 5.4). Overall, scores on the answers to the questions presented are a bit lower. The number of no responses and no strategies shows a slight increase, particularly in relation to the task involving sorting data. This might be due to the fact that in Exercises 1, the question regarding sorting data requires a single fact retrieval (*"Which is the country with the lowest death rate?"*), whereas the parallel question in Exercises 2 requires more than one fact retrieval (*"Which are the three most populated countries?"*). The task requiring the retrieval of objects with a given characteristic also shows an increase in the number of no responses and no strategies. Children seem to find it a bit more difficult to answer the question *"How many countries belong to Europe?"* than *"In how many countries is the population smaller than 3 million?"* . They are apparently equally difficult, although one includes numerical data and the other deals with alphanumerical data. Most students chose the option "same as" to answer question of level 1.2, but a few used "includes", obtaining 15 European countries instead of 14 (USSR was recorded as being European and Asian at the same time). Task level 1.1 *("In which continent is Benin?")* also shows an increase in the number of no strategies as well as in the number of wrong strategies. Common errors were the wrong selection of the field to search, as for example:

Search data
CONTINENT
includes
Benin

and the wrong description of the commands used. Two students used "Search" instead of "Display", without noticing that one can only search for a field at a time, although any number of fields can be displayed, and that a search requires the entry of a value or condition. One student's response to this question was:

Start a new search
Search data
COUNTRY
CONTINENT
Display

		EXERCISES 1						EXERCISES 2					
	TASK	RESPONSE			STRATEGY			RESPONSE			STRATEGY		
	LEVEL	R	W	N	R	W	N	R	W	N	R	W	N
RESPONSES	1.1	24	0	1	22	1	2	24	0	1	16	4	5
	1.2	22	3	0	23	1	1	19	0	6	19	0	6
	1.3	14	5	6	6	12	7	8	12	5	10	11	4
	1.4	25	0	0	22	3	0	21	2	2	19	1	5
	2.1	14	4	7	13	8	4	15	4	6	22	0	3
	Σ	99	12	14	86	25	14	87	18	20	86	16	23

	TASK	STRATEGIES				STRATEGIES			
	LEVEL	A	B	C	N	A	B	C	N
TYPES OF STRATEGIES	1.1	0	16	7	2	0	8	12	5
	1.2	0	5	18	2	0	5	14	6
	1.3	0	4	14	7	0	7	14	4
	1.4	0	11	14	0	2	7	11	5
	2.1	0	11	10	4	0	10	12	3
	Σ	0	47	63	15	2	37	63	23

Table 5.4 Overview of the results of the two sets of data retrieval tasks: Exercises 1 and 2: 13 year olds

The question requiring the retrieval of objects with more than one given characteristic (*"How many countries have a birth rate under 15 per 1000 and a death rate under 11 per 1000?"*) seems again to be the hardest for children to answer. The number of correct responses to this question decreased substantially in comparison to those from Exercises 1, but the opposite ocurred for the number of correct strategies employed. The types of error shown this time are again the same. Out of the ten students who used the strategy wrongly, seven did not link the two conditions logically, one used OR instead of AND, another one made a wrong selection of the fields to search (POPDOCTOR instead of BIRTHR) and the last one focused just on one variable. Again, only 4 students did a complete search correctly. A quite common incorrect response was:

Search data *BIRTHR* *less than* *15*	*Search data* *DEATHR* *less than* *11*
(15 out of 65 records)	*(37 out of 65 records)*

Regarding question number 3, concerning the finding of a relationship between two variables, this time more students (22 out of 25 compared to 13 out of 25 from Exercises 1) specified the strategies employed correctly. This may be due to their better understanding of the names of the fields they had to choose. In the first set of exercises they had to select GNP and URBANPOP (URBANPOP was not made explicit in the question). In the second set they had to choose BIRTHR and DEATHR, and both terms were made explicit in the sentence. However out of the 22 who used a correct strategy, only 15 gave a correct answer to the question. It seems that after having displayed the fields BIRTHR and DEATHR (strategy B), children had difficulty in deciding whether there was or there was not a relationship between the two variables. For some children, the same happened after having displayed a Scattergraph (strategy C). A few more students used strategy C correctly this time. Table 5.5 below shows the frequency of right and wrong uses of strategies B and C in both Exercises 1 and 2.

		EXERCISES 1					EXERCISES 2		
		Strategies					Strategies		
		B	C	N			B	C	N
	R	8	5			R	10	12	
Strategy	W	4	4		Strategy	W			
	N			4		N			3

Table 5.5 Relationship between correctness and nature of the strategies used for level 2.1 in Exercises 1 and 2: 13 year olds

The frequencies of each strategy used to solve the second sheet of data base tasks are quite similar to those of the first set of exercises, except for the questions at task level 1.1. In the first set of exercises, most students used strategy B, and in the second, most used strategy C, even though, as said before, the two questions are apparently quite parallel (*"What is the population of Japan"* and *"In which continent is Benin"* respectively).

The least efficient strategy, strategy A, was only used by two students, and then only for the level 1.4. Most of them again employed strategies B and C, strategy B being slightly less common than in Exercises 1.

Three students used strategy B and four strategy C solely in both sets of Exercises. This suggests that the type of strategy used, may be more task (or question) oriented than cognitively oriented.

5.2.2 Older students

Older students' performance on the tasks was quite similar to that of the 13 year olds. As did the younger ones, almost all older students gave a correct answer to the questions concerning the retrieval of a characteristic of an object, the retrieval of objects with a given characteristic (levels 1.1 and 1.2 respectively in the "logical" hierarchy of tasks), and also the question which required sorting data (level 1.4) in both Exercises 1 and Exercises 2. However, the number of no responses decreased substantially compared to the younger group. Table 5.6 shows an overview of older students' performance on the two sets of data retrieval tasks.

		EXERCISES 1						EXERCISES 2					
	TASK	RESPONSE			STRATEGY			RESPONSE			STRATEGY		
	LEVEL	R	W	N	R	W	N	R	W	N	R	W	N
RESPONSES	1.1	24	0	0	22	0	2	23	0	1	20	2	2
	1.2	23	1	0	24	0	0	22	2	0	21	2	1
	1.3	7	15	2	8	13	3	16	8	0	16	5	3
	1.4	23	1	0	23	1	0	24	0	0	23	0	1
	2.1	20	3	1	19	3	2	20	4	0	23	0	1
	Σ	97	20	3	96	17	7	105	14	1	103	9	8
	TASK	STRATEGIES						STRATEGIES					
	LEVEL	A	B	C	N			A	B	C	N		
TYPES OF STRATEGIES	1.1	0	14	8	2			0	15	7	2		
	1.2	0	8	16	0			0	12	11	1		
	1.3	0	7	14	3			0	6	15	3		
	1.4	0	8	16	0			0	7	16	1		
	2.1	0	10	12	2			0	9	14	1		
	Σ	0	47	66	7			0	49	63	8		

Table 5.6 Overview of the results of the two sets of data base tasks: Exercises 1 and Exercises 2: 16 year olds

The question requiring the retrieval of objects with more than one given characteristic (level 1.3) seems again the hardest for children to answer. In relation to the first set of exercises, just about one third of the students (seven) answered this question correctly (fourteen 13 year olds had answered this item correctly), and eight of them specified a correct strategy (seven younger students did). For the question *"Which are the African countries whose growth of population is higher than 2.5?"* , the most common type of error was the attention paid just to one variable or condition, like

Sort data
POPGROWTH
Display records
COUNTRY
POPGROWTH
Display

where children are using their knowledge of the domain (they know which the African countries are) to answer the question, and the use of OR as though AND, like

Search data
CONTINENT
same as
(Africa)
Add to the search (OR)
POPGROWTH
Higher than
(2.5)
Display these records
CONTINENT
POPGROWTH
Display

Unlike the youngsters, no one chose the wrong field to search (three younger pupils did) or failed to display the information after the search had been pursued.

In the second set of exercises, the number of correct responses increased substantially in comparison to those from Exercises 1. The number of correct strategies also increased. The types of error shown this time are again the same as in Exercises 1. Out of 6 students who used the strategy wrongly, 4 used OR instead of AND, and one focused just on one variable, and one mixed the commands (did not understand the meaning of Sort and Search). The number of no responses decreased slightly compared to that of the younger group.

Regarding task level 2.1 (finding of a relationship between two characteristics), the older students performed substantially better than the younger pupils in both sets of exercises. It is at this level where differences between the two groups become more evident (see Table 5.7). The selection of a wrong field to search was not common this time (one older student against seven younger students). Almost all students could give a correct answer to the question after having displayed a scattergraph, which suggests their possible better understanding of graphical representations.

13 year olds																			
EXERCISES 1										EXERCISES 2									
		Strategies					Strategies					Strategies					Strategies		
		B	C	N			B	C	N			B	C	N			B	C	N
	R	10	4			R	8	5			R	5	10			R	10	12	
Resp.	W	1	2	1	Strat.	W	4	4		Resp.	W	3			Strat.	W			
	N	1	3	3		N			4		N	2	2	3		N			3
16 year olds																			
EXERCISES 1										EXERCISES 2									
		Strategies					Strategies					Strategies					Strategies		
		B	C	N			B	C	N			B	C	N			B	C	N
	R	9	10			R	9	10			R	7	12	1		R	9	14	
Resp.	W	1	2	1	Strat.	W	1	2		Resp.	W	1	1	2	Strat.	W			
	N			1		N			2		N					N			1

Table 5.7 Responses and types of strategies for task level 2.1 - Finding of a relationship between two characteristics - Exercises 1 and 2: 13 and 16 year olds

Regarding the types of strategies employed to solve the problems, no one used strategy A in either set of exercises, and again strategy C was a little more common than strategy B. No important differences between Exercises 1 and 2 were found regarding types of strategies used, except for level 1.2, for which most students used strategy C in the first set of exercises, and B and C with the same frequency in the second set. Most younger students had used strategy C to answer the questions concerning the retrieval of objects with a given characteristic in both sets of exercises.

The older students did not show more consistency in the use of a particular strategy than the younger pupils. Four used only strategy C and three strategy B in Exercises 1. Again, four used only strategy C and three strategy B in Exercises 2, but just two were constant in both exercises (using strategy B).

5.2.3 Some conclusions

Overall, the results show a good success rate, the most difficult tasks being those which demand working with more than one variable at a time. According to the "logical" hierarchy of tasks, these are the retrieval of objects with more than one given characteristic (level 1.3) and the finding of a relationship between two variables (level 2.1).

The results indicate that the computer is not always used to its full advantage (strategy C) but there is a tendency to use the computer to do part of the work and leave some work for the children to do (strategy B). The latter implies the use of simpler commands (Display instead of Search) and an understanding of the computer as able to store data and show it in the form of lists, leaving for the user the tasks of checking and counting, but not an understanding of the computer as able to do logical calculations. It is good, however, that in practice strategy A (not exploiting the data base facilities at all) is not commonly used.

It seems that the use of strategy B or C depends to a certain extent on the demands of the task. So that most 13 and 16 year olds used strategy B for the tasks which require the retrieval of a characteristic of an object (level 1.1 - one fact retrieval), C for tasks 1.2, 1.3 and 1.4 (several facts retrieval and sorting), and strategies B and C indiscriminately for task 2.1 (relationship). The choice of one or another strategy does not seem to be so much influenced by the students' general perception of the search system, since only very few used the same strategy consistently along the different tasks.

5.3 Analysis of the results of the "report editing" task

The text of the report task is to be found in Chapter 3 (Fig. 3.2), with the statements to be tested indicated. A list of the ten statements included in the report, organised according to the "logical" hierarchy of tasks, can also be found in Chapter 3 (Fig. 3.3).

Three aspects were considered in the analysis of the report task. First, the number and type of statements children checked; second, the correctness of the responses, not only in terms of True and False statements, but also in terms of the paths they used to reach their answers; and third, the types of strategies used when interacting with the computer.

The analysis of the report task was done according to the same categories used for the analysis of the two sets of data retrieval exercises, both for the responses (Right, Wrong, No response) and for the types of strategies employed (strategies A, B or C).

5.3.1 Detecting statements of the report

Concerning children's ability to find questions that could be asked about the text, most of them detected most statements included in the report, except statement number 9 which was detected only by one of twenty four 13 year olds and eight out of twenty four 16 year olds. Statement number 4 was tested by only two thirds of the students (see Table 5.8).

Statement Number	Hierarchy Level	13 year olds	16 year olds
1	1.2	21	23
2	1.1	22	22
3	1.1b(*)	21	21
4	1.4	16	16
5	1.3	24	23
6	2.1	23	21
7	1.4b	22	23
8	1.3b	21	21
9	1.2b	1	8
10	2.1b	19	21

Table 5.8 Frequency of statements checked in the "report editing" task by 13 and 16 year olds
(*) 'b' means that children were facing a task of that particular level for the second time

An explanation of why statement 9 was not recognised as such could derive from an analysis of the organisation of the sentence. Statement 9, in the text, was *"Less developed areas, such as the three African countries Sudan, Uganda and Bangladesh, have many interesting features"*. The proposition to be tested is that all the three countries are in Africa. It is clear however, that the primary message of the sentence is not this (false) fact, but is *"Less developed countries have many interesting features"*, which is barely testable with a data base (if at all). The anaphoric subsidiary clause, *"such as the three African countries Sudan, Uganda and Bangladesh"* is, in the context of such a sentence, information treated as "given", not "new" (see Halliday, pp. 175-179). Thus it is very understandable why this statement was missed.

Statement 4 *"Its population is already the largest of the world"* , may have been missed because "India" is implicit in the sentence, and substituted by "its", which again is anaphoric.

Some other statements which were not intended to be checked, were checked by a few students (see Table 5.9).

STATEMENT	13 year olds	16 year olds
"The population of the world is very unevenly distributed"	3	6
"Besides, the world population is not expanding at the same rate everywhere"	2	4
"The gross national product, that is, the average income of a country, gives an idea of its development"	1	2
"How well off countries are is reflected by the GNP"	0	2
"In some countries, the need to reduce the population growth rates becomes urgent, and the reduction of birth rates seem to be the only means of doing so"	0	2
TOTAL	6	16

Table 5.9 Frequency of statements within the report, not intended to be checked, checked by 13 and 16 year olds

Though the statements in Table 5.9 can hardly be answered by simply retrieving data from a data base, two of them *"The population of the world is very unevenly distributed"* and *"Besides, the world population is not expanding at the same rate everywhere"* could to a certain extent be tested by displaying the fields COUNTRY, POPULATION and POPGROWTH for all the countries. However, the statement *"How well off countries are is reflected by the GNP"* could only be tested by children's use of their previous knowledge about which the well off countries are. The last two statements *"The gross national product, that is, the average income of a country, gives an idea of its development"'* and *"In some countries, the need to reduce the population growth rates becomes urgent, and the reduction of birth rates* *seems to be the only means of doing so*'" are explanatory. They provide additional information. They are not amenable to data base analysis.

The reason why more 16 year olds than 13 year olds checked this kind of statement could be explained in several ways: by closer attention to the text or by having more knowledge which suggests possible tests.

In the two sets of data retrieval tasks children had to deal only with the analysis of sentences. In the "report editing" task, however, children are dealing with the analysis of a text whose meaning is more than the sum of the meanings of its individual sentences. To understand a text, children have to be able to make links between sentence meanings, often over a considerable distance. This involves the resolution of anaphora, as seen above. It also involves a good deal of inferencing, observed when children called upon their knowledge of the world to answer particular statements

5.3.2 Responses and strategies

Going on to the analysis of the responses to the statements checked, the results (see Table 5.10) show that most 13 and 16 year olds did well on the statements involving the retrieval of a characteristic of an object (level 1.1), the retrieval of objects with a given characteristic (level 1.2), and sorting data (level 1.4). They not only gave a correct True / False answer but they also described the strategies employed correctly. Statement 9 (level 1.2b) was excluded from the analysis because it was only perceived as such by one 13 year old and eight 16 year olds.

		13 year olds						16 year olds					
Task	Stat.	Response			Strategy			Response			Strategy		
level	No	R	W	N	R	W	N	R	W	N	R	W	N
1.1	2	21	1	0	22	0	0	21	1	0	20	2	0
1.1b	3	20	0	1	21	0	0	19	2	0	19	2	0
1.2	1	21	0	0	21	0	0	21	2	0	21	2	0
1.3	5	11	13	0	10	14	0	20	3	0	18	5	0
1.3b	8	12	8	1	11	10	0	17	4	0	16	5	0
1.4	4	16	0	0	16	0	0	16	0	0	16	0	0
1.4b	7	22	0	0	22	0	0	23	0	0	23	0	0
2.1	6	11	12	0	11	11	1	13	8	0	7	14	0
2.1b	10	7	7	5	14	5	5	15	6	0	19	2	0
	Σ	141	41	7	148	40	6	165	26	3	159	32	0

Table 5.10 Overview of the results of "report editing" task : 13 and 16 year olds

Task levels 1.3 and 2.1 which require the combination of two variables at a time and use of logical condition and relation respectively are again the most difficult, especially for the younger pupils. Regarding the two statements involving the retrieval of objects with more than one given characteristic, less than half of the 13 year olds answered the statements correctly and specified a correct strategy, whereas more than two thirds of the older students did. More than half of the 13 year old students failed in the description of the strategies used, the most common type of error being the focus just on one of the variables. For example, to statement 5 *"The birth rate is over 50 in more than half of the African countries"*, a common type of answer using strategy C (partial search) was:

Search data
CBIRTHR
greater than 50
Display records
COUNTRY
CBIRTHR
Display

in which the variable CONTINENT has not been taken into consideration. An example of the same type of error but when using strategy B would be:

Display records
COUNTRY
CBIRTHR
Display

The older students performed better in tasks of level 1.3 than the younger ones, although only about one third (9 in task 1.3 and 7 in task 1.3b) used a complete C strategy, in which the logical operator AND is used to narrow down the search. Partial searches and strategy B were common (see Table 5.11).

The results show that in general 13 and 16 year olds' performance in the report editing task is very similar except for the tasks involving the retrieval of objects with more than one given characteristic, and for statement 10 *"In countries with low GNP, the infant mortality rates are lower than in rich countries"* (level 2.1b) which are better performed by the older group. Although more than half of the 13 year olds described a correct strategy to test the relationship in statement 10, less than one third of them answered the item correctly, and 5 did not even give an answer to the truth or falsity of the statement. This could be due to the complexity of the comparative sentence which in a sense deals with a double negation *("infant mortalities are lower than")*. The opposite happened, however, regarding 16 year olds' performance on statement 6, also a relationship (level 2.1), *"In European countries, for instance, the birth rates are very low which may be because there is a high percentage of women using contraceptives"*. Only 7 described a correct strategy, but 13 gave a correct answer. Whether this was due to their difficulty in describing the strategy on paper or to their ability to answer the question without using the data base is not clear.

The most common types of errors in the description of the strategies used to answer statements of level 2.1 (finding of a relationship) were: looking at the two characteristics separately and focusing on only one variable. Statement 6 *"While in*

European countries, for instance, the birth rates are very low which may be because there is a high percentage of women using contraceptives" was subdivided into two separate statements by some students. It was not uncommon to find answers to *"in European countries the birth rates are very low"* and *"in European countries there is a high percentage of women using contraceptives"* separately. Some other students chose the wrong field to search. For example, regarding statement 10 (see above) two 16 year old students chose the field BIRTHR instead of INFANTMORT. It seems that some children found it difficult to deal with a relationship that involves a negative term, such as infant mortality, and chose instead what they wrongly thought was its positive opposite.

Overall, again relationships and complex (AND / OR) searches appear to be the most difficult, but easier for the older students than for the younger ones. This difference in performance between the two age groups is more apparent in the results of the "report editing" task than in the results of the two sets of data retrieval tasks (Exercises 1 and 2).

In relation to the nature of the strategies used (see Table 5.11), no one used strategy A, and strategy C was a little more common than strategy B.

As said before, to answer statements which require the retrieval of objects with more than one given characteristic, quite a number of students made partial searches, which means that not many used the most efficient strategy by narrowing down the search with the use of the logical operator AND. Again, this suggests difficulties with logical operators.

		13 year olds		16 year olds	
Task level	Statement Number	Strategies B	C(*)	Strategies B	C
1.1	2	6	16	10	12
1.1b	3	10	11	11	10
1.2	1	9	12	12	11
1.3	5	8	16(9)	6	17(8)
1.3b	8	10	11(5)	9	12(5)
1.4	4	3	13	4	12
1.4b	7	10	12	8	15
2.1	6	15	8	14	7
2.1b	10	9	10	10	11
	Σ	76	109	84	107

Table 5.11 Frequency of the types of strategies used in the "report editing" task by 13 and 16 year olds
(*) The number in brackets under strategy C represents the number of partial searches

For most statements, strategy C was slightly more commonly used than strategy B. However, as an exception, substantially more students used strategy B to answer statement 6 (level 2.1) than strategy C. This could have something to do with the fact that such a statement involves a causal relationship, and children may not find the scattergraph representation appropriate for cause - effect relationships.

5.4 Is there a structure of data base tasks?

The results described in previous sections show that some tasks (levels 1.3 and 2.1) are more difficult than others (levels 1.1, 1.2 and 1.4) for both age groups, and that this difference in difficulty between the two classes of tasks decreases with age (see

Figure 5.1). (Note that scores of responses and strategies for Exercises 1 and 2 have been added one to the other. Tasks with a (b) correspond to Exercises 2).

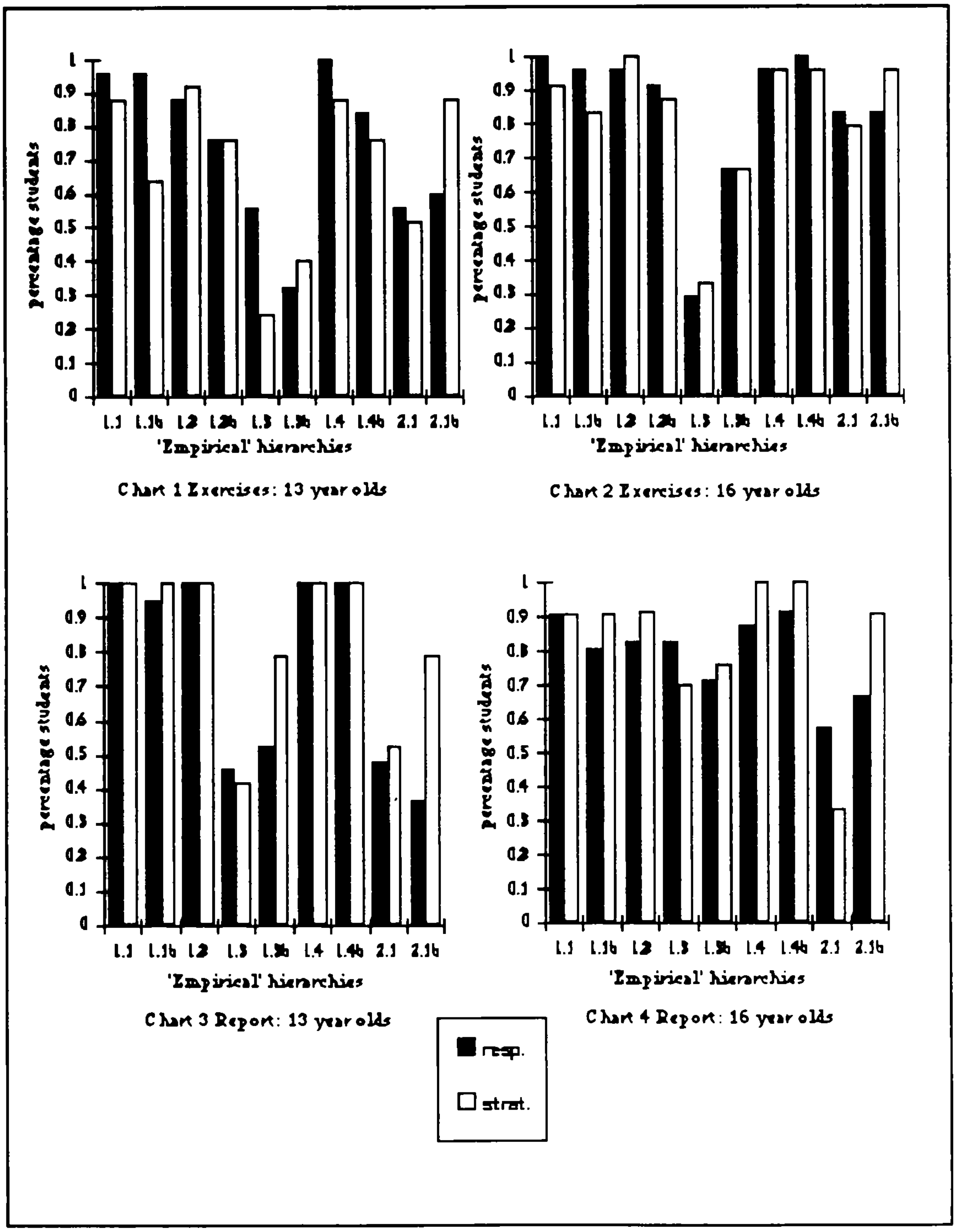


Fig. 5.1 Percentages of correct responses and strategies for Exercises 1 and 2 and "report editing" task: 13 and 16 year olds

Table 5.12 shows the "empirical" hierarchies of data base tasks obtained from the exercises and report for the 13 and 16 year olds. Percentages were calculated to avoid misinterpretation of the report results. So then, for task 1.4, a score of 16 does not mean that 8 were wrong on that task, but that 100 % got it right because only 16 students checked that particular statement. This means that sorting data appears to be in itself an easy data base task, but that it becomes rather difficult when it is embedded within the "report editing" task.

Exercises								Report (*)							
13 year olds				16 year olds				13 year olds				16 year olds			
Resp.	Aver.	Strat.	Aver.	Resp.	Aver.	Strat.	Aver.	Resp.	Aver.	Strat.	Aver.	Resp.	Aver.	Strat.	Aver.
1.4	1	1.2	.92	1.1	1	1.2	1	1.4b	1	1.4b	1	1.4b	.913	1.4b	1
1.1	.96	1.1	.88	1.4b	1	1.4	.958	**1.1**	1	1.1	1	1.1	.909	**1.4**	1
1.1b	.96	1.4	.88	1.1b	.958	1.4b	.958	1.2	1	1.2	1	**1.4**	.875	1.2	.913
1.2	.88	2.1b	.88	1.4	.958	2.1b	.958	**1.4**	1	1.1b	1	1.2	.826	1.1	.909
1.4b	.84	1.4b	.76	1.2	.958	1.1	.916	1.1b	.95	1.4	1	1.3	.826	1.1b	.904
1.2b	.76	1.2b	.76	1.2b	.916	1.2b	.875	**1.3b**	.523	2.1b	.789	1.1b	.809	2.1b	.904
2.1b	.6	1.1b	.64	2.1b	.833	1.1b	.833	2.1	.478	**1.3b**	.523	1.3b	.714	**1.3b**	.761
2.1	.52	2.1	.52	2.1	.833	2.1	.791	1.3	.458	2.1	.521	2.1b	.666	1.3	.702
1.3	.56	1.3b	.4	1.3b	.666	1.3b	.666	2.1b	.368	1.3	.416	2.1	.571	2.1	.333
1.3b	.32	1.3	.24	1.3	.291	1.3	.333								

Table 5.12 "Empirical" hierarchies of data base tasks: 13 and 16 year olds
(*) Numbers in bold correspond to the levels of the hierarchy which change position when percentages are calculated

Overall the data supports some hierarchy of difficulty of tasks, but not in the detail originally proposed. There appear to be two classes of tasks:

1. easy tasks, accomplished by the great majority of children at both ages:
    - 1.1 the retrieval of a characteristic of an object
    - 1.2 the retrieval of objects with a given characteristic
    - 1.4 sorting data on a given characteristic

2. harder tasks, which show noticeable improvement with age:

    1.3 the retrieval of objects with more than one given characteristic

    2.1 the finding of a relationship between two characteristics

In essentially every case (exercises and report), tasks 1.1 and 1.2 proved easy. The case of 1.4 is less clear. It must be noted, however, that all sorting tasks asked for one (or a few) extreme cases only, not using all the power of sorting facilities. This seems likely to have made the task rather easy, so it should not be assumed that any sorting task would fall in the first group. A reason why one sorting task (statement 4 in the "report editing" task) was unusually difficult has been suggested on page 109. Scores on these three tasks are very similar, and sensitive to specific questions / statements, and therefore it is difficult to say that they form a hierarchy.

The second group involves operations which are known to be relatively difficult. In chapter 6, it will be seen that the construction of logical sentences is particularly hard. Much evidence (eg. from work on proportion) indicates that seeing relationships between variables is difficult. However, by 16, more than half of students are succeeding at all these more difficult tasks (except at level 1.3 in the first set of exercises, at which only one third succeed), whereas at 13 the success rate varied from one third to a half (average scores varied from 28 to 60 %).

The detailed nature of tasks has, however, some effect, as for example level 1.3 being harder than level 2.1 in the exercises but not in the report; and level 1.4 being especially easy in the report but less so in the exercises. Even within the same task level, scores were sensitive to specific questions or statements. For example, twenty two 13 year olds described a correct strategy for task level 2.1 in Exercises 2, but only thirteen did so in Exercises 1. In the same way, nineteen 16 year olds described a correct strategy for task level 2.1b in the "report editing" task, but only seven did so for task 2.1.

This means there is little sense in trying to use these data to investigate the hierarchy in any great detail. In fact, the rank correlations between the original "logical" order and the ordering for the responses are all positive, and most are positive for the strategies (see Table 5.13).

	Log
Log	1
E13R	0.43
E13S	0.212
E16R	0.406
E16S	-0.03
R13R	0.433
R13S	0.333
R16R	0.417
R16S	0.167

Table 5.13 Spearman Rank Correlation of hierarchies ("empirical" and "logical")

Correlations for the responses are higher than those for the strategies. This may suggest that searching a data base and describing how to search a data base are two quite different activities, and that operations which are easy to perform are not necessarily easy to describe.

5.4.1 Structure and consistency

The reliability or internal consistency (KR20) of the various data base tasks varied essentially from 0.5 to 0.72 (see Table 5.14). These moderate values could (but need not) indicate a complex factor structure. With few items per task (9 or 10), and relatively few subjects per task, little reliance can be placed on a factor analysis. Since the results of factor analysis are however slightly suggestive, they are briefly reported here (detailed factor tables can be found in Appendix 5).

Exercises				Report			
13 year olds		16 year olds		13 year olds		16 year olds	
Response	Strategy	Response	Strategy	Response	Strategy	Response	Strategy
0.575	0.654	0.282	0.638	0.721	0.583	0.5	0.639

Table 5.14 Reliability coefficients (Kuder Richardson 20) for data retrieval tasks and "report editing" task: 13 and 16 year olds

The Bartlett sphericity test indicated the presence of some structure. In separate factor analysis of each task (exercises and report), there were between 3 and 5 factors for a task. There was some tendency for broadly similar factors to appear. One factor could usually be associated with simple data base operations (for the younger students, the theoretically simplest data base operation (one fact retrieval) appears as a unique factor), and another one with a general data base ability, involving most levels of the hierarchy. Another factor, often appeared to relate to logical operations and relations, though not always in a clear cut way, sometimes dividing between these and sometimes involving sorting as well (maybe because sorting directly and the testing of relationships indirectly share the characteristic of being prompted by the menu options SORT and GRAPHS and FIGURES). However, it does not seem wise to rely at all heavily on such results. They derive from correlations which are necessarily restricted in range because several tasks were very easy, and because the data were restricted in range (0, 1 or 2). At best, we see some suggestion of support for the division of the hierarchy into simple operations versus logical operations and relations.

5.5 Some conclusions

Overall, it appears that data base analysis is complex and difficult to operationalise. Performance on data base tasks is not only affected by the level of difficulty of the task, but also by the peculiarities of the particular question asked. Even within the same level, questions do not share points in common. Performance on data base tasks is influenced by the nature of the task in which they are embedded. What might be easy in the two sets of data retrieval tasks (search directed questions) might be difficult in the "report editing" task (non search directed plus text based). Statement 4 in the "report editing" task (sorting data) is an example. Questions and text analysis and children's content knowledge may be important aspects to consider in data base analysis. However, a clear distinction between two levels of difficulty of tasks emerges, with one type of task (1.3) logically assigned a fairly low position in a hierarchy of difficulty, in fact belonging to the more difficult level.

Chapter 6 CHILDREN'S UNDERSTANDING OF LOGICAL OPERATORS

6.1 Analysis of the results of the Interpreting and the Constructing tasks

As described in Chapter 3, two logical tasks were designed and given to the students, one about interpreting search sentences involving logical operations, and one about constructing such sentences. The analysis of the constructing and interpreting tasks will be discussed in this chapter (both questionnaires can be found in Appendix 3). The younger students will be discussed first, followed by the older pupils. It will be convenient to present first the results from the pre-test, and then later to make comparison with those from the post-test. The results of the two questionnaires will first be presented separately, comparisons between them being made later.

Twenty five 13 and twenty five 16 year olds completed the interpreting task in the pre-test. Twenty five 13 year olds did it in the post test as well, but only twenty 16 year olds responded to the task in the post-test. Twenty five and twenty four 13 and 16 year olds respectively completed the constructing task in both pre- and post-tests.

A useful feature of these tasks is that the responses are non-verbal. They consist just of choosing shelves in a diagram which are those chosen by the logical sentence (Interpreting task) or circling those components of a logical sentence presented in sets of alternatives which allow to get the chosen shelves (Constructing task).

The results of these two tasks for the younger group have been published (Bezanilla and Ogborn, 1992) and are therefore reported here fairly briefly. The paper is bound into the thesis as Appendix 1.

6.1.1 The interpretation of a logical sentence as a pattern of search

The results of the interpreting task were analysed according to the following criteria:

- number of correct answers
- number of wrong answers, including:
    - -interpretation of AND as though OR and/or OR as though AND.
    - -interpretation of the logical sentence as one of the statements separately, either 'p' or 'q', but not as the logical relation between them.
    - - other and no responses

For the purpose of the analysis, the eight items or logical sentences were grouped in three groups: the tautology and contradiction, the three other sentences using AND, and the three other sentences using OR.

6.1.1.1 Younger students

Results from this first questionnaire on children's understanding of the logical operators AND and OR are shown in Table 6.1. The results are rather clear cut and show several interesting features.

To start with, the AND sentences were much better understood than the OR sentences. Almost all students gave a correct answer to those items whose logical forms are (p & q), (~p & q), and (~p & ~q). Though some experimental research shows that people respond more slowly and uncertainly to negative than to affirmative statements (Wason, 1962; Evans, 1972), students here seemed not to have problems in coping with negative AND statements.

	13 year olds			
AND	Correct Answer	As though OR	'p' or 'q' separately	Other/No Response
p & q	23	0	2	0
~p & q	24	0	1	0
~p & ~q	23	0	1	1
OR	Correct Answer	As though AND	'p' or 'q' separately	Other/No Response
p v q	15	3	7	0
~p v q	0	16	6	3
~p v ~q	0	22	3	0
Contradic./ Tautology	Correct Answer	As though OR/AND	'p' or 'q' separately	Other/No Response
p & ~p	0	21	4	0
p v ~p	8	0	17	0

Table 6.1 Responses to Interpreting task: pre-test: 13 year olds

In relation to the OR sentences, more than half of the students gave a correct answer to the item having the logical form (p v q), in which both statements are affirmative. However, the logical forms (~p v q) and (~p v ~q) seemed to be interpreted generally as (~p & q) and (~p & ~q) respectively. OR was understood as though AND. This special interpretational difficulty of negative OR statements could be due to the fact that in natural language disjunctions are rarely formulated with negative components (Evans, 1972). Evans shows that on exclusive OR problems, (~p v ~q) is easier than (~p v q), but on inclusive problems this difference in difficulty is not evident. The results of this study support these findings. Neither the double-negative sentence nor the single-negative one was interpreted correctly by any of the students. Another interesting result is that OR sentences were more often interpreted as just one of the statements (usually 'p') than were AND sentences (16 against 4). This could be explained by the supposed ambiguity of the inclusive and exclusive senses of OR.

Children seemed to experience difficulty in interpreting contradictions (p &~p) and tautologies (p v ~p). With regard to the contradiction *"Goods which are cameras and which are not cameras"*, most students ticked all the shelves instead of none, thereby converting the contradiction into a possible proposition (the tautology). (p v ~p) was generally interpreted as 'p'. In this case, OR was understood exclusively by children, as if it indicated the choice of one alternative. Thus, their interpretation of the tautological sentence *"Goods which are TVs or which are not TVs"* led most children to tick just the TV row.

6.1.1.2 Older students

The results of the older pupils show a quite similar pattern to those of the younger group. However, they present several distinct features which will be now discussed.

Again most students answered the AND statements correctly, although a bit worse than the 13 year olds (see Table 6.2). Surprisingly, 7 students interpreted statement (p & q) as (p v q). In one secondary school (school C) the order of the statements presented was altered by mistake. Page 3 was presented before page 2. This means that the first item children answered was a complex one, the double OR negation *("Goods which are not CD players OR which are not Japanese")* instead of a simple one of the form (p & q) *("Goods which are CD players AND which are Japanese")*. Whether this had any influence in the results is not clear.

	16 year olds			
AND	Correct Answer	As though OR	'p' or 'q' separately	Other/No Response
p & q	17	7	1	0
~p & q	22	1	1	1
~p & ~q	20	1	3	1
OR	Correct Answer	As though AND	'p' or 'q' separately	Other/No Response
p v q	20	0	5	0
~p v q	2	5	13	5
~p v ~q	1	20	4	0
Contradic./ Tautology	Correct Answer	As though OR/AND	'p' or 'q' separately	Other/No Response
p & ~p	0	23	2	0
p v ~p	14	0	11	0

Table 6.2 Responses to Interpreting task: pre-test: 16 year olds

Regarding OR statements, most 16 year olds (83 %) answered the (p v q) statement correctly (against 60 % of the 13 year olds). Scores for this item were even higher than for (p & q) and equal to (~p & ~q). However, the older pupils also seem to have problems with negative OR statements. Whereas most 13 year olds interpreted (~p v q) as (p & q), most 16 year olds interpreted it as 'p' or 'q' separately. Single negative OR statements are understood exclusively by the 16 year olds. (~p v ~q) was understood as (~p & ~q). Like the youngsters, they understood OR as though AND. As said before, the fact that in natural language it is uncommon to use negative OR statements may explain these findings.

Contradiction and tautology appeared to be very difficult for the older students as well. A few more gave a correct answer to the tautology this time, and about half interpreted it as 'p'. The contradiction (p & ~p) was again understood as the tautology.

6.1.2 The building up of a logical sentence to achieve a desired pattern of search

The Constructing task was analysed by registering the logical forms used by the students to achieve the presented patterns of search. In this way we can not only see whether their responses are right or wrong, but also detect possible common errors. As before, for ease of discussion, the younger pupils will be discussed first, followed by the older pupils.

6.1.2.1 Younger students

Table 6.3 shows the frequency of each type of logical sentence for each pattern of search. Bolded figures on the diagonal indicate the number of correct answers, that is, a perfect match between the structure of data presented and the logical sentence constructed by the students.

Answer / Question	p & q	~p & q	~p & ~q	p v q	~p v q	~p v ~q	p & ~p	p v ~p	Other/ No Resp.
p & q	**19**	4							2
~p & q	6	**13**	2						4
~p & ~q	6	7	**6**	3		1			2
p v q	13	1	2	**4**					5
~p v q	9	8 (2)*	2	2	**1**				3
~p v ~q	7	7	7 (1)			**1** (1)			3
p & ~p			20			1	**0**		4
p v ~p	12	1	1	7				**0**	4
Σ	72	41	40	16	1	3	0	0	27

Table 6.3 Responses to Constructing task - pre-test: 13 year olds
(*) The number in brackets indicates the number of students who constructed a sentence of that particular form, but with a different 'p' or 'q'. An example will be provided below.

From Table 6.3 it can be seen that AND sentences were much more often and better constructed than OR sentences. Most students answered the item regarding the logical form (p & q) correctly and more than half of them did it in the case of (~p & q). However, they were less successful in building up those sentences which required the use of OR. Almost none gave a correct answer to the items (p v q), (~p v q), (~p v ~q). They used AND to join the statements, instead of OR, and they usually constructed the sentences in affirmative form (p & q).

The patterns of search corresponding to both of the two double negative forms (~p & ~q) and (~p v ~q) were translated equally as (p & q), (~p & q) and (~p & ~q) (AND sentences). The questionnaire did not allow the expression of (~p v ~q) as (~(p & q)), nor of (~p & ~q) as (~(p v q)). Two out of the eight students who constructed (~p & q) sentences to achieve a pattern of the form (~p v q) chose a wrong "field" to search. As an example, to define the pattern of search shown in Fig. 6.1 (~p v q), one of these students wrote: *"Are not Scottish AND are men"* (~p & q), instead of *"Are not men AND are English"* (also ~p & q).

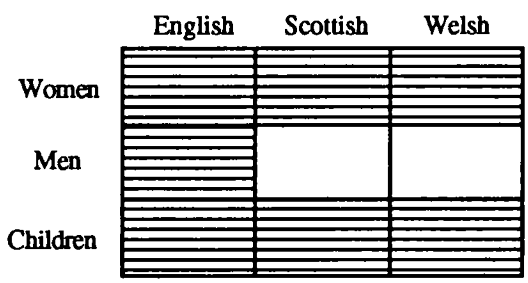


Fig. 6.1 Pattern of search of the form (~p v q) - Constructive task

As will be seen later on, this type of error was also common in the post-test, and even more common in the responses of the 16 year olds.

Another interesting result is that the structure of data corresponding to the form (p v q) was translated mainly as (p & q). The treatment of OR as AND is present here also.

No correct responses were given for the contradiction (p & ~p) and the tautology (p v ~p). Also, no one constructed any sentence of these forms. The pattern representing the form (p & ~p) (empty grid) was basically expressed as (~p & ~q) by essentially all students. To describe a structure of data with no data, students seem to have seen this as 'not p' and also 'not q'. It is as if children associated the emptiness of the shelves with the logical double negation. To achieve the pattern of search corresponding to the tautology (p v ~p) (all shelves are filled), most children constructed sentences of the logical forms (p & q) and (p v q), in which the two statements are affirmative.

Overall, children scored much better in the Interpretation task than in the Constructing task. The task of building up a structure of data, given a logical sentence, seems to be easier than the task of building up a logical sentence, given a particular structure of data. In each case, AND sentences are better understood and constructed than OR sentences. Although in the interpreting task, AND sentences (p & q), (~p & q) and (~p & ~q) are equally well interpreted, children had more difficulty in constructing negative AND sentences than positive ones in the constructing task. Negative OR sentences seem to be extremely difficult both to interpret and construct.

Regarding the contradiction (p & ~p), it is interesting to notice that although in the first questionnaire it is interpreted as the tautology (p v ~p), when having to build the logical sentence in the second questionnaire, students produce the form (~p & ~q). The tautology (p v ~p) is interpreted mainly as 'p' and constructed as both (p & q) and (p v q).

As was expected, the number of no responses was higher in the constructing task than in the interpreting task.

6.1.2.2 Older students

The responses of the older students to the constructing task can be seen in Table 6.4. In general, the pattern of responses is quite similar to that of the youngsters, except for the pattern of search showing double AND negation (~p & ~q) which is three times better achieved by the 16 year olds than by the 13 year olds. Also, the pattern of search corresponding to the form (p v q) was slightly better translated by the 16 year olds. Overall performance is relatively higher (61 perfect matches between pattern of search and logical sentences by the older pupils against 43 by the younger pupils).

Answer / Question	p & q	~p & q	~p & ~q	p v q	~p v q	~p v ~q	p & ~p	p v ~p	Other/ No Resp
p & q	**21**	1	1						1
~p & q	4	**16** (1)							4
~p & ~q	2	1	**17**		1	1			2
p v q	9 (2)		2	**7**					6
~p v q	4	10 (5) *	3		**0**				7
~p v ~q	2	3	13			**0**			6
p & ~p	1	2	7	1		4	**1**		8
p v ~p	12	1		4				**0**	7
Σ	55	34	43	12	1	5	1	0	43

Table 6.4 Responses to Constructing task: pre-test: 16 year olds
(*) The number in brackets indicates the number of students who constructed a sentence of that particular form, but with a different 'p' or 'q'.

AND sentences were again more often and better constructed than OR sentences. The pattern of search equivalent to the logical form (~p v q) was mostly translated as (~p & q), but again out of the 10 students who wrote a (~p & q) sentence, 5 chose either a wrong 'p' or a wrong 'q'. Some examples (see Fig. 6.1 above) are: *"Are Scottish and are not men", "are Welsh and are not men", "are English and are not men", "are English and are not children", "are men and are not Scottish".*

Whereas the pattern representing the form (~p v ~q) was equally interpreted as (p & q), (~p & q) and (~p & ~q) by the 13 year olds, more than half of the older students interpreted it as (~p & ~q). Children's difficulty to distinguish between the logical operators AND and OR is evident here as well.

Children failed in the construction of those sentences corresponding to the pattern of search involving contradiction (empty cells) and tautology (all shelves are filled). Unlike the younger pupils, who mostly translated the (p & ~p) pattern as (~p & ~q), less than one third of the older pupils used such a logical form. Another third either did not answer this item or constructed a logical sentence of the forms (~p & ~p) or (~p v ~p), in which only one characteristic (nationality or types of people) is chosen and double negation is applied. These responses are: *"Are not English and are not Scottish"*, *"are not Scottish and are not English"*, *"are not English or are not Scottish"*, *"are not women and are not men"*, *"are not men and are not women"* . More than half of the students (66 %) used in any case double negation, as if they associated it with the emptiness of the shelves. Obviously, if the nationalities had been presented in one list and the types of people in the other, instead of having both in each list, this type of response (~p & ~p), (p & p), (p v ~p), etc. would have been avoided. The reason for the presentation chosen was that in building a search sentence using a data base, all choices or fields to search are in fact available at every selection.

Children also failed in the construction of tautologies. Half of them used the form (p & q) and almost one third either did not answer the item or gave an unexpected answer of the kind (p & p) or (p v p). Some examples are: *"Are English and are English"*, *"are English and are Scottish"*, *"are English and are Welsh"*, *"are men and are women"*. In opposition to the contradiction, most responses (87 %) to the tautology were affirmative.

Overall, children did better in the interpreting task than in the constructing task, but unlike the younger students who had difficulties in the construction of negative AND sentences, they could interpret and construct negative AND sentences relatively well. Therefore, the difference of performance between the two tasks, regarding AND sentences, is not so evident for the 16 year olds. Affirmative OR sentences are three times easier to interpret than to construct. Negative OR sentences, as for the younger pupils, are very difficult both to interpret and construct.

The tautology was interpreted and constructed as either (p & q) or 'p' (or the equivalent (p & p) in the constructing task), but the contradiction was interpreted as the tautology (p v ~p) (AND was understood as AND AS WELL) and constructed as either (~p & ~q) or (~p & ~p).

6.1.3 Pre-test - Post-test Comparisons

This section analyses to what extent children's understanding of the logical operators AND and OR shows any improvement after having taught them how to interrogate a data file with GRASS, and having worked with GRASS to solve a number of data base tasks of different levels of difficulty. Chapter 5 shows that, after having taught them the difference between AND and OR, children still experience some difficulties when solving those tasks which require the retrieval of objects with more than one given characteristic. Therefore, much improvement is not expected from the post-test. The results of the interpreting task will be discussed first, followed by those of the constructing task.

6.1.3.1 The Interpreting task

Table 6.5 shows there is no important difference between the pre and post-test scores on the interpreting task for both 13 and 16 year olds. The results from the post-test show there was not only no improvement in performance, but a slight worsening in the case of the younger pupils, particularly regarding the forms (~p & ~q), (p v q) and the tautology (p v ~p).

13 year olds										
Pre	Post	p & q	~p & q	~p & ~q	p v q	~p v q	~p v ~q	p & ~p	p v ~p	Σ
✓	✓	21	20	18	7	0	0	0	1	67
X	X	0	0	1	6	21	24	25	17	94
✓	X	2	4	5	7	0	0	0	7	25
X	✓	2	1	1	4	4	1	0	0	13
16 year olds										
Pre	Post	p & q	~p & q	~p & ~q	p v q	~p v q	~p v ~q	p & ~p	p v ~p	Σ
✓	✓	9	13	16	9	0	0	0	9	56
X	X	4	2	0	2	16	20	19	14	67
✓	X	4	4	1	7	1	0	0	1	18
X	✓	3	1	3	2	3	0	1	6	19

Table 6.5 Comparison of pre- and post- tests for Interpreting task: 13 and 16 year olds

The analysis of the younger pupils shows that for the AND items (excluding contradiction), scores on both pre-test and post-test were high. Slightly more children did worse on the post-test than on the pre-test, for the two items using negation. In the OR items (excluding tautology), essentially no children were correct in either test on the items involving negation. For (p v q), the majority were correct on one or both occasions. Of those who were wrong on one occasion rather more were wrong on the post-test. On the contradiction and tautology, most were wrong on both occasions. Since the main result is that essentially nearly all students did very well on AND items and very badly on all OR items except (p v q), there is little point in testing statistically for any changes.

Why they did a bit worse in the post-test than in the pre-test is not clear. Obviously, working with GRASS did not help them to better understand the operators AND and

OR. It is true, though, that using GRASS does not require the interpretation but the building up of logical sentences.

The analysis of the older pupils shows that about half of the students answered correctly the affirmative items (p & q) and (p v q), and the tautology (p v ~p) in both pre-test and post-test. Two thirds answered correctly item (~p & q) and three fourths did (~p & ~q). Surprisingly, negative AND sentences (excluding contradiction) were better interpreted than the affirmative one. Almost no one did well on negative OR items, excluding the tautology. Overall, there was no improvement, but unlike the younger group, there was no worsening.

6.1.3.2 The Constructing task

Regarding the 13 year olds, Table 6.6 shows the results for the constructing task, with pre-test responses at the top left of each cell and post-test responses on the bottom right.

For the AND items (excluding contradiction), there is a slight improvement on those items involving negation (~p & q) and (~p & ~q). The logical form (p & q) gets the highest score on both pre-test and post-test. For the OR sentences (excluding tautology), scores were equally low in the pre-test and post-test. On the contradiction and tautology, all students but one were wrong on both occasions. It is interesting to notice that on the pre-test, the pattern representing the contradiction (p & ~p) was generally expressed as (~p & ~q). However, on the post-test, only half of them expressed it in such a form. The other half either did not answer that item or gave a response of the kind (~p & ~p) and (~p v ~p), as the older students did. AND sentences were more often constructed than OR sentences on both pre-test and post-test. However, the results from the post-test show that there is a small increase in the number of students who constructed negative OR sentences, although without success. The number of no responses to the negative OR sentences, contradiction and tautology was slightly higher on the post-test than in the pre-test.

Question \ Answer	p & q	~p & q	~p & ~q	p v q	~p v q	~p v ~q	p & ~p	p v ~p	Other/No Response
p & q	19 / 18	4 / 2			/ 1	/ 1			2 / 3
~p & q	6 / 3	13 / 16	2 / 2		/ 1				4 / 3
~p & ~q	6 / 1	7 / 5	6 / 11	3 / 2	/ 1	1 / 2			2 / 3
p v q	13 / 15	1 / 3	2 / 1	4 / 3					5 / 3
~p v q	9 / 10	8 / 5	2 / 1	2 / 1	1 / 1				3 / 7
~p v ~q	7 / 5	7 / 7	7 / 5		/ 2	1 / 1			3 / 5
p & ~p			20 / 10			1 / 2	0 / 1	/ 2	4 / 10
p v ~p	12 / 11	1 /	1 / 1	7 / 4		/ 1	/ 1	0 / 0	4 / 7
Σ	72 / 63	41 / 38	40 / 31	16 / 10	1 / 6	3 / 7	0 / 2	0 / 2	27 / 40

Table 6.6 Comparison of pre- and post-tests for Constructing task: 13 year olds (pre-test above diagonal; post-test below diagonal)

In relation to the older students, the results show a slight improvement on the items involving single negation (~p & q) and (~p v q), and, in opposition to the 13 year olds, a slight worsening on the items (p & q), (~p & ~q) and (p v q). Even so, the older students still outperform the younger ones in the AND items, and in the affirmative OR (see Table 6.7).

Question \ Answer	p & q	~p & q	~p & ~q	p v q	~p v q	~p v ~q	p & ~p	p v ~p	Other
p & q	21 / 18	1 / 1	1 / 1		/ 1	/ 1			1 /
~p & q	4 / 1	16 / 18	/ 1		/ 1				4 / 1
~p & ~q	2 / 1	1 /	17 / 15		1 /	1 / 1			2 / 5
p v q	9 / 15		2 / 1	7 / 5					6 / 1
~p v q	4 / 5	10 / 9	3 / 2		0 / 2				6 / 4
~p v ~q	2 / 1	3 / 7	13 / 8			0 / 0			6 / 6
p & ~p	1 /	2 /	7 / 6	1 /		4 / 1	1 / 1		8 / 14
p v ~p	12 / 4	1 / 1	/ 2	4 / 1				0 / 1	7 / 13
Σ	55 / 45	34 / 36	43 / 34	12 / 6	1 / 3	5 / 2	1 / 1	0 / 1	43 / 44

Table 6.7 Comparison of pre- and post-tests for Constructing task: 16 year olds (pre-test above diagonal; post-test below diagonal)

More than two thirds of the students constructed a (p & q) statement for the pattern of search corresponding to the form (p v q) in the post-test, whereas just above one third had done it in the pre-test.

The pattern of search representing the tautology was not mainly constructed as (p & q), as in the pre-test, but as (p v p) and (p & p). Actually, three students constructed sentences of the form (p v p), and five of the form (p & p). For the contradiction, seven constructed sentences of the form (~p and ~p) and four of the form (~p v ~p) in the post-test, whereas in the pre-test, only five children used any of these univariable forms. The analysis of the post-test then shows a tendency for children to interpret the empty cells (contradiction) and the fully filled cells (tautology) as 'not p' and 'p' respectively.

6.2 Some conclusions

The present results show that children experience difficulty in interpreting and constructing logical sentences involving AND, OR and NOT, particularly negative OR sentences, tautologies and contradictions. Even after having explained the use of AND and OR as logical operators, using examples, and having given children the opportunity to solve some data base tasks, most of them continued to have difficulties of the same kind as before with these operators. It should be noted that GRASS expresses AND as "narrow down the search" and OR as "add to the search", thus giving its own interpretation of the logic.

To what extent and in what way this may be due to linguistic problems in understanding the forms AND, OR, NOT, or to the nature of the mental operations required to interpret and construct logical sentences, is not clear. Some evidence points towards a linguistic interpretation, notably results indicating the use of inclusive OR as if it were exclusive OR, and the use of OR as AND when negation is involved. According to some interpretations of Piaget, children who have attained the stage of formal operations should be able to work with propositional reasoning, and therefore, to understand and construct logical AND and OR sentences. However, children's understanding of language connectives seem to influence their performance on reasoning tasks. (Conjunctions are the most frequently occurring connectives. Disjunctions are much more rarely formulated and heard by children). To what extent and in what way linguistic competence plays a significant role in the acquisition of advanced logical thinking is left open by Piaget.

The analysis of the contradiction and tautology deserves special attention. These logical forms are true or false by virtue of their logical form. Osherson and Markman (1975) call them "non-empirical" because their truth value does not depend upon states of the world (a fact of course exploited by Wittgenstein in his *Tractatus Logico-Philosophicus*). To understand them, children need to examine the form of the sentence itself, which presupposes that children can use language not only as a code for reality but also as an object of thought in its own right. Children must be able to regard language independently of the reality it refers to. In their study, Osherson and Markman emphasise the role of "meta-linguistic" skills rather than linguistic competence per se in evaluating contradictions and tautologies. The evidence here presented points to the great difficulty of doing so.

The task requiring constructing of a logical sentence proved much more difficult than the task which required interpreting such a sentence, especially for the 13 year olds, despite the care taken to give them parallel forms. It can be reported informally that the subjects did not complain of the difficulty of either task, and completed both quickly, and easily. One explanation of this result could be that it is linguistic in origin. AND almost completely replaces OR (even in (p v q)), perhaps used in the sense "AND AS WELL" (AND as logical addition, that is to say, OR). Another explanation could be the need in the sentence construction task to see the pattern of filled and empty cells as a whole, and then to convert it into a logical sentence, by contrast with the interpretation task in which cells can be ticked one at a time. These explanations deserve some attention in further research.

Overall, the results indicate that constructing logical data base queries is not a task to be taken for granted. One may expect difficulties to arise mainly in the use of negation and of OR (search widening) queries. AND queries (search narrowing) seem on this evidence to be by far the most straightforward. It is fortunate that in practice non-negative AND queries are likely to be common. Since in natural language negative OR sentences, contradiction and tautology are rarely used, it is not expected that children will often formulate questions of that kind when using a data base.

The present results also bring out one of the main problems of using natural language for the construction of queries. As Shneiderman (1980) has stated:

> *"By allowing users to use natural language without training we allow the ambiguities of English syntax to pollute the query process, driving developers to design long and tedious clarification dialogs" (Shneiderman, 1980, page 208)*

Chapter 7 CHILDREN'S UNDERSTANDING OF THE NATURE OF DATA

7.1 Analysis of the results

The purpose of the questionnaire about children's understanding of the types of information that can be contained in data bases is discussed in detail in Chapter 3, section 3.2.3. The questionnaire itself can be found in Appendix 3, section D. The results for the younger students will be presented first, followed by those of the older students. Results from the pre-test will be discussed first, followed by results from the post-test.

The results of this questionnaire were analysed in terms of the number of correct and incorrect answers, taking account of different types of questions. Twenty five 13 year olds and twenty 16 year olds responded to the questionnaire in both pre- and post-tests.

As described in chapter 3, the questionnaire had twenty items of different types. Each gave a question. The student did not have to answer the question, but to say whether the question could, in principle, be answered from a suitable data base, or not. These questions are:

• questions which can be answered from a data base, involving:

- retrieval of one or several objects:
  - Question 1: *"Who is the Prime Minister of the British Government?"*
  - Question 4: *"How many cars were bought in England in 1988?"*
  - Question 6: *"Which is the title of the first book written by Agatha Christie?*
- sorting data:
  - Question 8: *"Which is the highest mountain in North America?"*
  - Question 12: *"Which is the biggest country of the EEC?"*
  - Question 19: *"Which was the warmest month in Britain last year?"*

- finding a relationship between two variables:

  Question 9: *"What would happen to Sweden's population if there was a sudden baby boom in the next five years?"*

  Question 14: *"What is the relationship between the production rate of grapes and the price of the wine?"*

  Question 17: *"Do richer countries have more people studying at Universities?"*

  Question 20: *"Is there a link between the number of inhabitants in a town and the number of schools?"*

• questions which can not be answered from a data base, related to:

- Future:

  Question 2: *"Will it rain more this year than it did this year?"*

  Question 10: *"Is John Major going to be reelected in the next elections?"*

- Judgment:

  Question 3: *"Which is the most objective British newspaper?"*

  Question 16: *"Which was the best directed film last year?"*

- Intention:

  Question 5: *"Is the Soviet Union serious about arms reductions?"*

  Question 13: *"Do football fans not really mean any harm?"*

- Feelings:

  Question 7: *"Do children enjoy using the computer?"*

  Question 15: *"Do girls like getting flowers as a present?"*

- Context:

  Question 11: *"Is it more convenient to go to Bristol by coach or by train?"*

  Question 18: *"Do people wear smart clothes at a party?"*

7.1.1 Younger students

The analysis of the results for the 13 year olds shows overall a good success rate. All students responded to all questions, but none answered all twenty questions correctly. However, 18 out of 25 students answered fifteen or more questions correctly.

The results of the younger group show that most children knew that those questions related to Future and Intention cannot be answered by any data base. Scores on questions about Feelings and Judgment were rather lower. Responses to the two questions under the category of Context were quite surprising in the sense that almost all students correctly said that question number 18 (*"Do people wear smart clothes at a party?"*) could not be answered by a data base, but only eight of them replied correctly in the same way for question number 11 (*"Is it more convenient to go to Bristol by coach or by train?"*). A reason for this may be that question 18 refers to people whereas question 11 is impersonal. Table 7.1 shows the average of correct answers to each of the types of question.

Type of question	Question No.	Average of correct responses	Total average
Future	2, 10	0.90	
Judgment	3, 16	0.66	
Intention	5, 13	0.88	0.76
Feelings	7, 15	0.74	
Context	11, 18	0.64	
DB - retrieval	1, 4, 6	0.73	
DB - sorting	8, 12, 19	0.85	0.73
DB - relationships	9, 14, 17, 20	0.6	

Table 7.1 Children's understanding of the nature of data: 13 year olds: pre-test

Scores for the questions involving relationships were also varied. Whereas most students gave a correct answer to questions 14 and 20 (*"What is the relationship between the production rate of grapes and the price of the wine?"* and *"Is there a link between the number of inhabitants in a town and the number of schools?"* respectively), very few answered questions 9 and 17 correctly. Question 9 (*"What would happen to Sweden's population if there was a sudden baby boom in the next* *five years?"*) is related to future. Therefore it is understandable that children thought it could not be answered by a data base.

The results show that students scored on the average higher on those questions which required sorting data than on those which required either the retrieval of one or several objects or the finding of a relationship, the latter getting the lowest scores. In some cases, there was little consistency in the responses to the questions grouped under the same category. For example, scores for answers saying that question 1: *"Who is the Prime Minister of the British Government"*' (retrieval of an item) can be answered by a data base were pretty low, but on the contrary, in responses to question 4: *"How many cars were bought in England in 1988"*' (retrieval of several items) nearly all saw it as answerable by a data base.

The same happened in relation to the questions which involve the finding of a relationship. Only ten students correctly said that question 17 (*"Do richer countries have more people studying at Universities?"*) could be answered by a data base. However, twenty gave the same (correct) response to question number 14 (*"What is the relationship between the production rate of grapes and the price of the wine?").*

It should be pointed out that such inconsistency of responses did not occur in the case of the three questions involving sorting data. Moreover, these got the highest scores.

Results from the post-test show a quite similar pattern to those from the pre-test regarding those questions related to constructs not amenable to data base analysis (Future, Context, Feelings, Judgment and Intention). A slight worsening is apparent though in questions concerning Future, Feelings and Context. However, students scored on the average higher when recognising questions which can be answered by a data base. An increase in the total average of correct answers is shown, moving from 0.73 in the pre-test to 0.81 in the post-test (see Table 7.2).

Type of question	Question No.	Average of correct responses	Total Average
Future	2, 10	0.84	
Judgment	3, 16	0.66	
Intention	5, 13	0.94	0.74
Feelings	7, 15	0.64	
Context	11, 18	0.62	
DB - retrieval	1, 4, 6	0.81	
DB - sorting	8, 12, 19	0.89	0.81
DB - relationships	9, 14, 17, 20	0.72	

Table 7.2 Children's understanding of the nature of data: 13 year olds: post-test

Again, the categories of Future and Intention are better understood as not able to be contained in a data base, than are the categories of Judgment, Feelings and Context. Inconsistency in the responses to the two Context questions is also apparent this time (see Fig. 7.1).

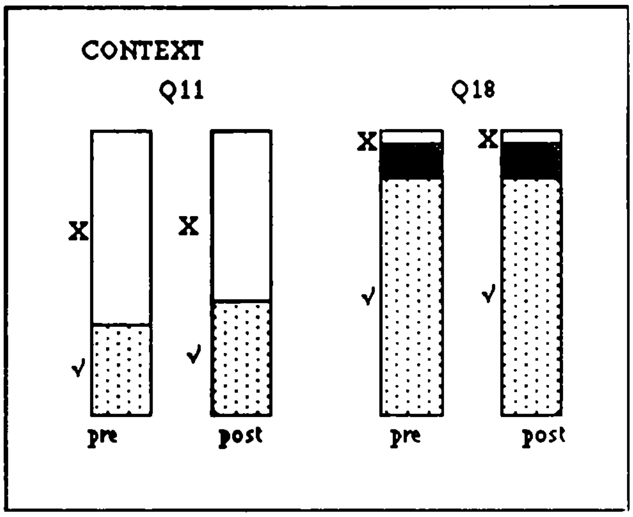


Fig. 7.1 Scores to questions 11 and 18 (Context): 13 year olds: pre- and post-tests
■ *represents the number of children who were correct in the pre-test and wrong in the post-test*

Substantial improvement was shown in children's understanding of the types of data that can be retrieved from a data base, particularly regarding the questions involving the retrieval of specific items of information and the finding of relationships. Again, questions involving sorting data got the highest scores in pre- and post-tests (see Fig. 7.2).

In figures 7.1 and 7.2 the dark shaded region represents children whose performance became worse, being correct in the pre-test but wrong in the post-test. Thus, these should be seen as belonging to the lower half of the bar diagram (correct) for the pre-test, but to the upper half (wrong) for the post-test.

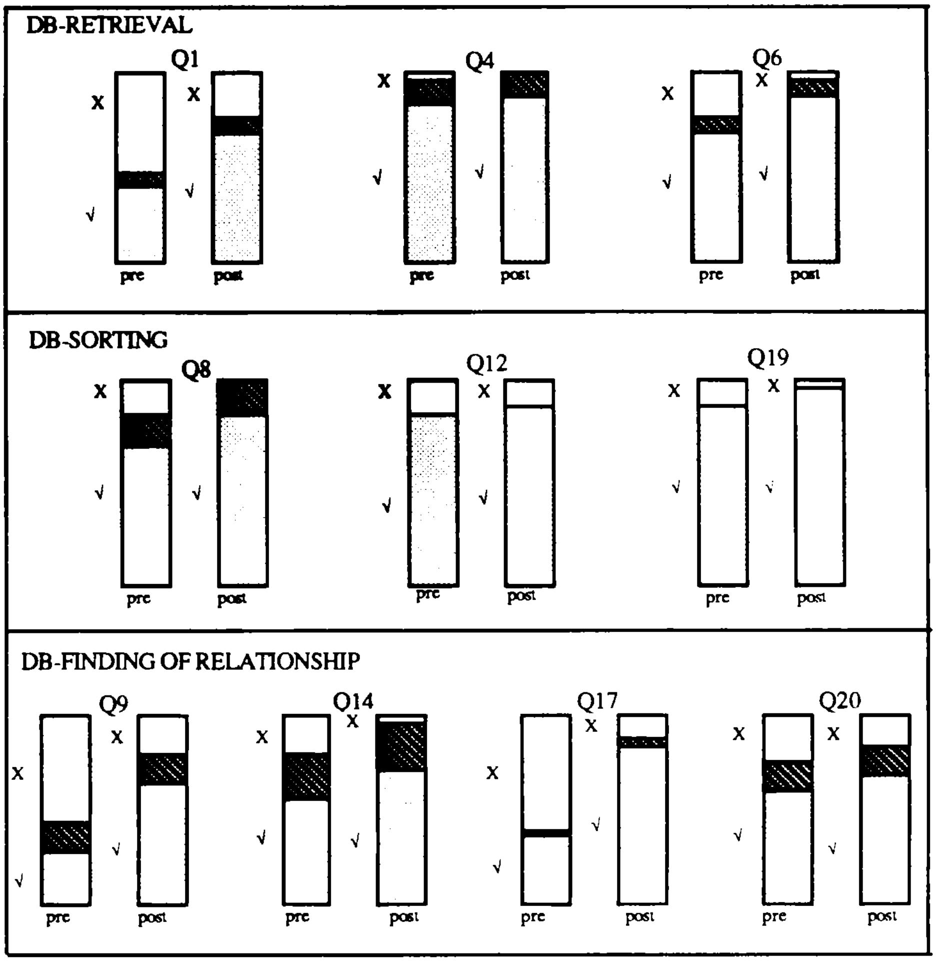


Fig. 7.2 Responses to the questions answerable by a data base: 13 year olds: pre- post-tests comparisons:
represents the number of children who were correct in the pre-test and wrong in the post-test

As can be seen from Fig. 7.2 scores for questions 9 *("What would happen to Sweden's population if there was a sudden baby boom in the next five years")* and 17 *("Do richer countries have more people studying at Universities?")*, both relationships, show the highest increase in the post-test. In fact, the use of a t - test indicates that difference between pre- and post-tests regarding questions answerable by a data base is just statistically significant ($t = 1.71$; $df = 24$; $p \leq 0.05$). The overall difference between the two tests, however, is not statistically significant ($t = 1.34$; $df = 24$; $p \leq 0.1$). Thus, we cannot confidently reject the hypothesis that children's understanding of the types of information that can be contained by data bases stays the same after having been trained in the use of GRASS. However, we can reasonably suppose that after having learned the use of GRASS and having performed a number of data base tasks of different levels of difficulty (including retrieval, sorting and testing of relationships) for a period of at least two hours, children's understanding of the types of questions which are answerable by a data base improves somewhat.

7.1.2 Older students

The results of the older students show overall a very good success rate in nearly all categories. All students answered all questions in pre- and post-tests. 80% or more of the pupils answered 15 (75 %) or more questions correctly in the pre-test and all did so in the post-test. Table 7.3 shows the average scores for each type of question in both pre- and post-tests. The older group outperformed the younger group in all types of questions and in both tests.

		Pre-test		Post-test	
Type of question	Question No.	Average of correct responses	Total Aver.	Average of correct responses	Total Aver.
Future	2, 10	0.9		0.95	
Judgment	3, 16	0.67		0.75	
Intention	5, 13	0.97	0.82	0.94	0.89
Feelings	7, 15	0.8		0.92	
Context	11, 18	0.75		0.9	
DB - retrieval	1, 4, 6	0.86		0.9	
DB - sorting	8, 12, 19	0.96	0.86	0.98	0.91
DB - relationships	9, 14, 17, 20	0.75		0.87	

Table 7.3 Children's understanding of the nature of data: 16 year olds: pre- and post-tests

Children scored higher on questions which can be answered by a data base than on those not amenable to data base analysis in pre- and post-tests, and within the first class of questions, again sorting was the data base operation that got the highest scores (all students answered correctly questions 12 and 19 in both pre- and post-tests). Also, as before, the testing of relationships got the lowest scores, though substantially higher than for the 13 year olds. In chapter 5, it was seen that performance on data base tasks involving the testing of relationships showed a noticeable improvement with age. These results support the view that 16 year old students understand data bases as able to deal with the finding of relationships between two variables substantially better than the 13 year olds. Fig. 7.3 highlights this difference between the two age groups.

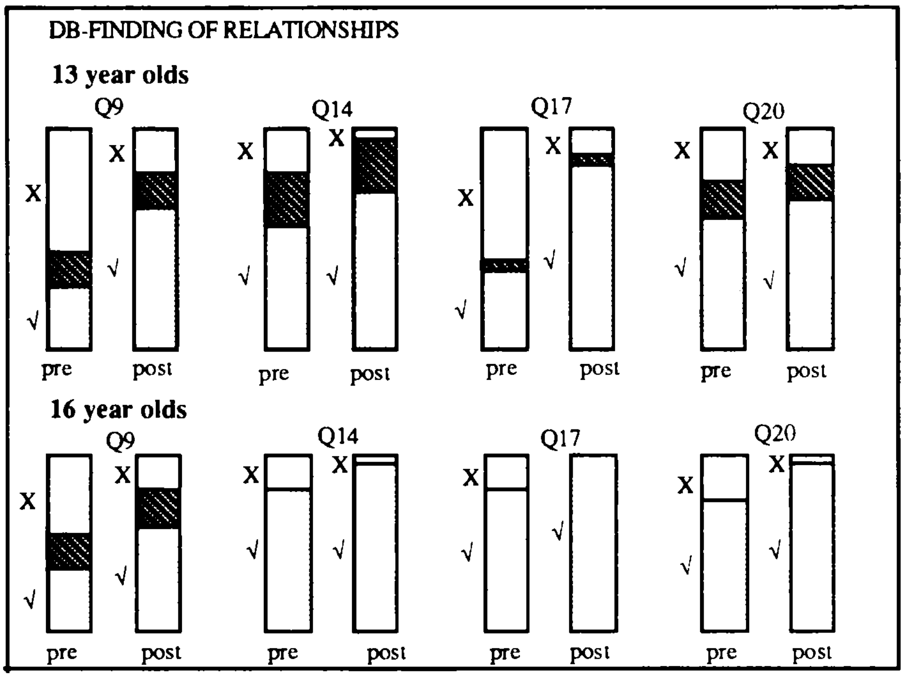


Fig. 7.3 Scores to questions involving relationships: 13 and 16 year olds: pre- and post-tests

Question 9 *("What would happen to Sweden's population if there was a sudden baby boom in the next five years")* is quite tricky (see pages 141-142), since it demands the prediction of population change in the future. Therefore, it is not surprising that it gets the lowest scores. It should be noted, however, that the POP data file used in the work with these students includes the field DOUBLETIME (time in years for the population to double in size if the present growth rate continues), which may facilitate its answer.

Questions concerning Future, Context, Intention and Feelings also show high scores. Judgment gets the lowest score. In all but one case, there is improvement in the post-test. Differences between the two age groups are particularly noticeable regarding the constructs Feelings, Judgment and Context (see Fig. 7.4).

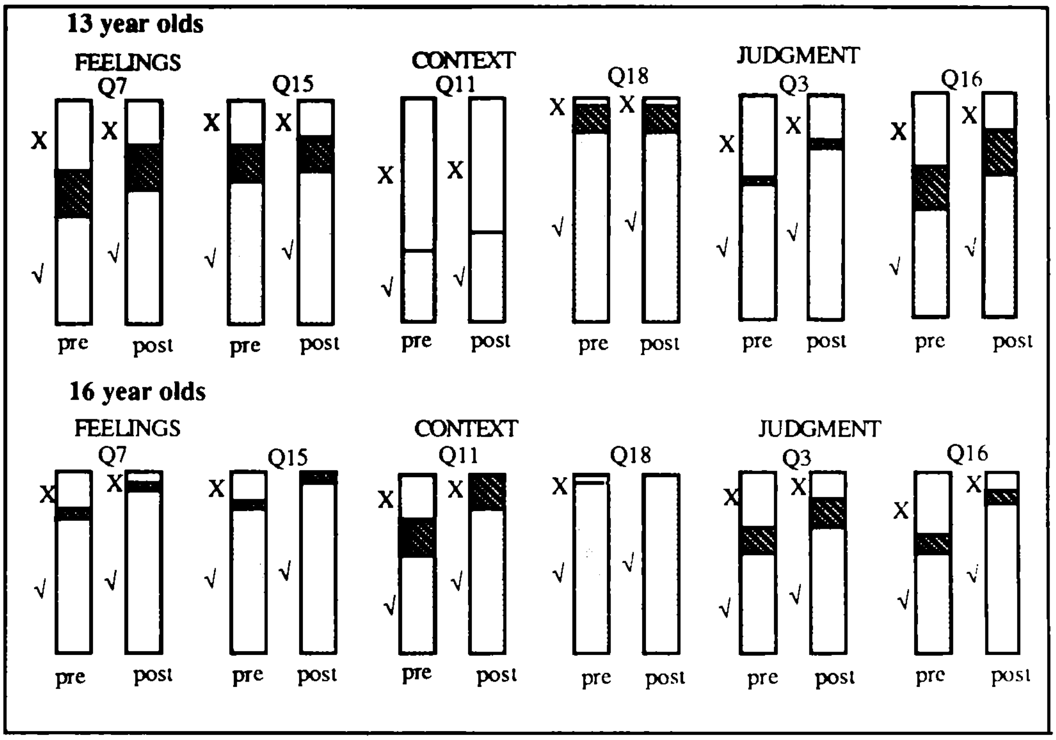


Fig. 7.4 Scores to questions related to the categories of Feelings, Context and Judgment: 13 and 16 year olds: pre- and post-tests

Unlike for the younger students, a t-test analysis related to the types of information that can be retrieved from a data base indicates that the difference between pre- and post-tests is not statistically significant ($t = 1.39$; $df = 19$; $p \leq 0.09$), but, the overall difference between pre- and post-tests is statistically significant at 0.05 level ($t = 2.52$; $df = 19$; $p \leq 0.01$).

7.2 Some conclusions

These results indicate that overall children have relatively little difficulty in distinguishing between the types of information that can be retrieved from a data base and the types of information that can not, with the older students being appreciably better than the younger ones. It seems that both 13 and 16 year olds can better discriminate what can be contained by a data base than what can not, and among the latter type of questions, they seem to have special difficulties with information concerning Judgment, Feelings and Context, particularly the younger pupils. However, only two questions under each category were presented, and results for a pair were not always very consistent. Thus the content of the question also has some influence.

In relation to the questions that can be answered by a data base, those involving sorting get the highest scores for both 13 and 16 year olds. Children do not hesitate to attribute to data bases the capability of sorting data. The retrieval of one or several items of information also gets high scores, whilst scores for the testing of relationships are a bit lower. Overall, the older students outperform the younger ones in the ten questions amenable to data base analysis. On questions which can not be answered by a data base, older children only do a shade better than the younger ones, but improve in the post-test, whereas the younger ones do not.

Overall, the general level of performance is good to excellent. At the pre-test, the younger pupils and the older ones are much the same for questions which cannot be answered from a data base, and at the post-test only the older ones improve. On questions which can be answered from a data base, the older students are better than the younger ones in both pre- and post-tests, and both improve.

Children seem to understand what the computer can and can not answer. In the next chapter, it will be seen whether the children can or cannot formulate meaningful and answerable questions, which is a rather different issue. Also, it will be shown how children understand data bases, not in terms of the nature and types of data that can be contained in them, but in terms of the nature of the data search and structure.

Chapter 8 CHILDREN'S UNDERSTANDING OF THE NATURE OF DATA SEARCH AND STRUCTURE

8.1 Introduction

As discussed in chapter 3, the questionnaire designed to test children's understanding of the ways information is kept, organised and found in a data base was arranged according to the "logical" hierarchy of tasks. Questions about how children imagined the computer could do some data base operations on imaginary data about childrens' books, at the five levels of the hierarchy, were presented. Thus, for example, question 1 *"Suppose we tell the computer to find the name of the author of a book titled 'Jane Eyre'. How do you imagine it might try to do it?"* corresponds to level 1.1 (one fact retrieval), and question 5 *"Suppose we tell the computer to find out if the most expensive books usually have the most pages. How do you imagine it might try to do it?"* corresponds to level 2.1 (the finding of a relationship between two variables).

The five specific data base tasks and the hierarchy levels they belong to can be seen in Fig. 8.1. The actual questionnaire can be found in Appendix 3, section E.

In the last part of the questionnaire, children were asked to write any other questions about the books which they thought the computer could answer. The purpose of this was first to see if children were capable of asking meaningful questions of a data base, and second to see at what level of the "logical" hierarchy of tasks the proposed question could be located.

Quest.No.	Hierarchy level		Data base tasks
1	1.1	Retrieval of a characteristic of an object	To find the name of the author of a book titled "Jane Eyre"
2	1.2	Retrieval of objects with a given characteristic	To find the titles of all the fairy tales
3	1.3	Retrieval of objects with more than one characteristic	To find all the textbooks which are cheaper than £5
4	1.4	Sorting data on a given characteristic	To put the books in order according to their price, ranging from the cheapest to the most expensive
5	2.1	Finding a relationship between two variables	To find out if the most expensive books usually have the most pages

Fig. 8.1 Data base operations children are asked to imagine the computer can do

In Chapter 7, it was seen that children seem to have an adequate model of the computer in terms of the types of information it can and cannot store. They are sure a data base can not deal with questions about Future and Intention, but they are not so sure it can deal with questions involving relationships, particularly the younger pupils. The questionnaire to be discussed in this chapter is an attempt to look a little more deeply at how children think data can be organised and found in the computer, which may reflect their models of how the computer works.

Children were tested before and after being trained in the use of GRASS, so that possible changes in their mental models of how information is organised and searched in a data base could be glimpsed. Twenty six 13 year olds responded to the questionnaire in the pre-test, and twenty five did so in the post-test. Twenty five 16 year olds answered the questionnaire in the pre-test, but only twenty did it in the post-test. The results for the younger students will be presented first, followed by those for the older ones. Results from the pre-test will be discussed first, followed by results from the post-test.

8.2 Method for the analysis of the results

The analysis of the results of this questionnaire was not easy. The children's responses were not always clear or explicit, and it was therefore difficult to develop a method to categorise them. It was my interest to know from their responses what their ideas were about how data is stored, linked and retrieved in a data base, and also to see the kind of language they used in their responses. Relevant features an answer could exhibit were defined. The relevant features, particularly those about the ways information could be found (strategy) were different for each task level (the same for the hierarchy levels 1.1, and 1.2). Thus, from each child's response we could identify some features of the way he thinks data is kept, organised and found in a data base, and then by looking at all the responses together, we could get some overview of children's understanding of data search and structure.

The feature analysis was done using a grid (see Fig. 8.2).

8.2.1 Feature analysis

Five aspects were considered in the analysis:

1. How data is stored
2. How data is linked
3. How data is found (search strategy)
4. What type of language is used
5. How efficient the strategy is

Feature Analysis	S1	S2	S3	S4	S5	....
How is data stored?						
just kept somewhere	✓	✓				
in an order			✓			
How is data linked?						
not linked	✓		✓			
somehow linked		✓				
record-field linked						
Strategy						
look for target directly	✓		✓			
look for known features		✓				
show – one field						
show – both						
What kind of language is used?						
ordinary language	✓	✓	✓			
data base query language						
Does method work?						
yes – well						
yes – inefficiently						
no	✓	✓	✓			

Fig. 8.2 Feature analysis for task levels 1.1 and 1.2 (simple retrieval)

The features for the analysis will be now explained in more detail. For this purpose, examples of the responses of the 13 year olds to question 1: *"How do you imagine the computer could find the name of the author of a book titled 'Jane Eyre'?"* (the retrieval of a characteristic of an object) will be provided.

1. How is data stored?

Most of children's responses did not refer particularly to how data is stored. It seems they think that data is just 'kept' somewhere in the computer (the computer just looks for the information until it finds it). The following responses to question 1 (level 1.1 in the hierarchy) would belong to this category:

> *"Go through to get list of information about Jane Eyre and read off the author"*
>
> *"It would search through the authors, then it would list all the books written by that author"*

A few students, however, thought that data is stored in a particular order (e.g. alphabetical), as is revealed in these responses:

> *"Go through in alphabetical order until you find it"*, or
>
> *"It will look up the letter E"*

2. How is data linked?

At one extreme, data could be seen as not linked at all (that is, that the computer just looks for what it needs), as the following responses suggest:

> *"It would go through the titles until it finds Jane Eyre in its memory"*
>
> *"By going through the cards and find the card that says name of the authors"*
>
> *"It would go through its memory until it found all the information about 'Jane Eyre'"*

At the other end, data could be record-field linked (you search for particular fields and particular records are matched). The only response under this category was:

> *"It will look up the field name 'title of the book' and search for 'Jane Eyre'. It will then look up the author and show it to you"*

Responses of the kind:

> *"Title of book*
> *Name of author/s*
> *Display information"*

which show an understanding of the computer as being able to show you linked columns of information (data in the same row belong to a same record) were also considered as belonging within the record-field linkage feature.

A rather common type of response involved an understanding of data as not being subject to perfect record-field linkages, but linked somehow. Often, children thought that by finding 'x', you could get 'y', as in the examples below:

> *"Go through all the cards until it gets to one with the title 'Jane Eyre'. Then see who the author is"*, or

> *"It would sort through the authors, then it would list all the books written by that author"*

Regarding question 4 (sorting data), some responses show data seen as not linked, like the following one:

> *"Go through all the cards and follow them in order"*

> *"Go through in alphabetical order and find the most expensive first"*

Where the response was of the kind "by sorting on prices, titles are re-arranged", it was labelled as "record-field linked". For example:

> *"Look at "price" and programme it to put them in order"*

> *"Sort through prices into order. Book titles on screen"*

> *"It would get the price data and somehow get them into the right order you told it"*

For question 5: *"How could the computer find if the most expensive books usually have the most pages"* (the testing of a relationship), responses suggest that data was seen as either not linked:

> *"Find out how many each book has and then one with the highest amount of pages is checked against the highest price and see if it is the same book"*

or "somehow linked":

> *"By looking at the books already in their order of price taking the most expensive ones and finding out how many pages they have, then do the same for the others"*

Data could also be seen as record-field linked, but only one student saw it that way in the pre-test:

> *"Compare the number of pages list with the order of price list and see if the hypothesis is correct"*

Example from the post-test:

> *"You'd sort data and have expensive books first. Then you'd ask it to display number of pages and price"*

3. How is data found? (Strategy)

For question 1 and 2 (simple retrieval tasks), students could say that the computer could just look for the target directly (look or list) or that it could look for known features (known data values for another field), or that it could just show the information contained in those particular fields (one or both) (see Fig. 8.2).

For question 1: *"How do you imagine the computer might try to find the name of the author of a book titled "Jane Eyre"?"*, the following responses would serve as examples of the above strategies:

- looking for the target (i.e. author) directly:

> *"By going through the cards and find the card that says name of the authors"*
>
> *"Search through the data banks of authors alphabetically"*

- looking for a known feature (i.e. "Jane Eyre" as the title of the book):

> *"It would go through its memory until it found all the information about 'Jane Eyre'"*
>
> *"It would search through the cards until it gets to 'Jane Eyre' and all the information will be there on it"*

• showing the information contained in one or two fields. None of the 13 year olds used this strategy in the pre-test. However, seven did in the post-test. The following examples are from the post-test:

> *"Display records*
> *Author of book*
> *Name of book"*

> *"By using Search Data it would search "Title of the Book" and "Name of author's", then it would display it and you would go through each book to find it"*

The two first categories described above would involve a Search option, whereas the last one would involve Display. These features were applied to the two first questions (the retrieval of a characteristic of an object and the retrieval of objects with a given characteristic). The analysis of strategy for the third question (the retrieval of objects with more than one given characteristic) was slightly different, as shown in Fig. 8.3.

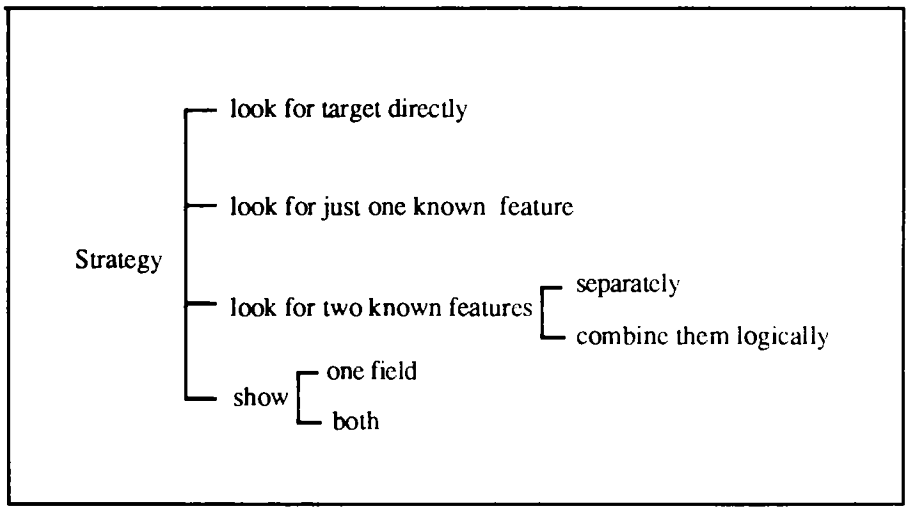


Fig. 8.3 Feature analysis (strategy) for task level 1.3 (the retrieval of objects with more than one given characteristic)

Responses to question 3: *"How do you imagine the computer could find all the textbooks which are cheaper than £5"* show that most students thought the computer would just focus on one variable, mainly price, as in the examples below:

> *"It would search under 'price' and the computer would see how much each book was and work out all the ones under £5"*

"It would look through the price list"

Some other students, however, thought that the computer would look for the two given features, but separately:

"Find all the textbooks, go through prices to see all the ones under £5"

None said that the computer would combine the two features logically (by the AND connector) in the pre-test, however two students did so in the post-test:

"Books under £5
AND
Kind of book textbook"

"Search data
price
less than £5
narrow down the search
kind of book"

Only one student , and only in the post-test, used the "show" strategy (data from one or both fields are displayed on the screen). In the example below, the strategy includes a search on "price" before the fields "price" and "kind of book" are displayed:

"Display all records
sort data
smaller than (price) £5
kind of book
display"

For those questions which did not involve retrieval, but sorting data and finding relationships between two variables, the strategies children used in their responses were very different to those used for the data retrieval questions. Regarding the question requiring sorting data *("How do you imagine the computer could put the books in order according to their price, ranging from the cheapest to the most expensive")* , some children thought the computer could sort data by simply looking at the prices, as can be seen in the following responses:

"Find the books price list"

"By looking at the book names list and price list and then showing out"

Some other children thought the computer could use an "extreme" strategy, as the follow examples show:

> *"Find the most expensive books and then go down"*

> *"The computer could find the cheapest book and the most expensive book, then finding the nearest to each book making two list, then add one to the other"*

Some children seem to take for granted that the computer can sort (re-arrange) data (the computer can just do it):

> *"Put each book in its memory and sort them logically"*

> *"It would go through all the cards and put them in order"*

Within this group, some students are more explicit when they refer to the "price" field specifically:

> *"Give some sort of 'sort' command in the price section"*

> *"Look at 'price' and programme it to do put them in order"*

For question 5 *"How do you imagine the computer could find if out if the most expensive books usually have the most pages"* (relationship), three strategies were distinguished (see Fig. 8.4).

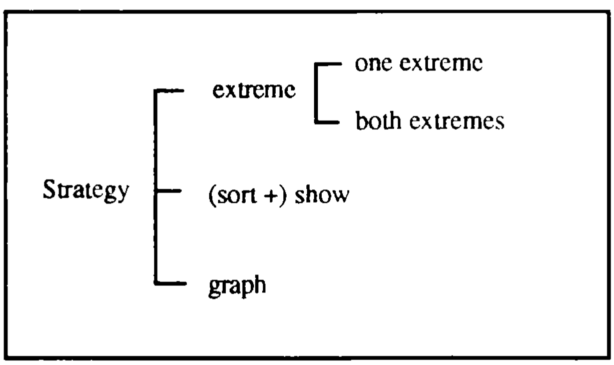


Fig. 8.4 Strategy for task level 2.1 (the finding of a relationship between two variables)

First, an "extreme" strategy. Relationships are inspected by looking at the high values for one variable (i.e. number of pages) and / or the other variable (i.e price), and / or to the relatively low values. Examples would be:

> *"Get the five least expensive and the five most expensive books and then get number of pages"*
>
> *"Make a list of all the expensive books, then see how many pages they have"*

In the first example, the two extremes are considered, whereas in the last example, just one extreme (the higher values) is looked at.

Second, a "show" strategy (displaying both variables simultaneously on the screen) or a "sort + show" strategy. This means sorting data on a characteristic (e.g. price) and then examining the values of price and the other characteristic (e.g. number of pages) by displaying the value of both variables. The following examples from the post-test illustrate the "show" strategy and "sort + show" strategy respectively, the latter being much more efficient than the former:

> *"Display records*
> *title of book*
> *number of pages*
> *price"*
>
> *"You would sort data and have expensive books first. Then you would ask it to display number of pages and price"*

Finally, a "graph" strategy. Some responses (mainly from the post-test) said that variables or characteristics could be related graphically by means of a scattergraph. Here are some examples:

> *"It would probably do it in graph form. Price on one axis and pages on another axis"*
>
> *"Draw a scattergraph of price against number of pages"*

This last strategy is the most efficient one, because when the number of records is large, it is not possible to examine the relationship between two characteristics by displaying all the values of the two variables on one screen ("sort + show" strategy).

4. What kind of language is used?

Here the features noted are whether children used ordinary language, or used parts of some data base query language, when explaining how to ask the computer for particular information. It was expected that in the post-test some children would employ the syntax of GRASS.

5. Does method work?

This refers to whether what children said could or could not be done by the computer, and if it could be done, whether efficiently or not, judging by how much work the method leaves for the user to do. Possible errors in the formulation of the queries were looked for, particularly from the post-test responses.

8.2.2 Difficulties of the analysis

The qualitative features are not easy to apply consistently, and there is sometimes doubt about the best features to characterise a given answer. Because the responses were often brief and inexplicit, it was usually easier to assign a lower rather than a higher level feature. This effect biases the results towards a lower rather than a higher performance. Where a higher level feature is scored, one can be fairly confident that it is present. This needs to be borne in mind in the following discussion of the results.

The difficulties are illustrated by the following response to question 3 *("How the computer could find all the textbooks which are cheaper than £5"):*

> *"The computer could search under all the types of books to find all the textbooks, then going through the prices finding all the ones under £5"*

This could be interpreted as follows: The computer finds the "kind of book" field and searches for the textbooks, then it narrows down the search with an AND option, finds the "price" field and searches for those textbooks whose price is less than £5.

But it could also be interpreted in another way: The computer finds the "kind of book" field and searches for the textbooks, then starts a new search, finds the "price" field and searches for all the books cheaper than £5. Thus it is difficult to decide whether the child means that the two features or characteristics are combined logically, or that they are not combined at all.

Regarding question 4: *"Could the computer put the books in order of price ranging from the cheapest to the most expensive?"* , could we infer from the type of response shown below that children think data is "record- field" linked (i.e. that by sorting on prices, book titles are re-arranged)?:

> *"Sorting the price smallest first"*
>
> *"Go through prices, picking out most expensive first, then go down to cheapest"*

In general, where there was doubt, the lower level of the possible features was coded for a given response.

8.3 Presentation of the results

As said previously, twenty six 13 year olds answered the questionnaire in the pre-test and twenty five did so in the post-test. Twenty five 16 year olds answered the questionnaire in the pre-test, but just twenty did in the post-test.

A few either did not answer one or more questions or gave an "un-understandable" response, or simply repeated the question asked. In the post-test, this kind of response was almost nil (see Table 8.1). Examples of "un-understandable" responses to questions 5 (relationship) and 1 (one fact retrieval) in the pre-test would be:

> *"Query Price Expensive*
> *Number of pages"*
>
> *"Type Query 'Jane Eyre' Name of author"*

The above responses are from a 16 year old, and they show a misconception concerning the way searches are done with a data base (i.e. the need to use the word

'Query' at the beginning of the sentence). In the post-test, this particular student answered questions 5 and 1 as follows:

"Display records
price and number of pages
display"

"Display
title of book
author
title of book"

This means he correctly assimilated the syntax of GRASS to his existing conceptions.

Other students' responses consisted of a mere enumeration of fields or characteristics, such as the following examples for questions 3 and 1 respectively:

"Price of books
Title of books"

"Title of books
Name of authors
Number of pages"

Since it was very difficult to get information about the way the students thought data could be kept, organised and found in a data base from these types of responses, these were grouped under an "uninterpretable" category.

Quite a number of students used the "naming fields" strategy, particularly in the post-test (see Table 8.1). This may suggest that children specified the characteristics they wanted to see together on the screen, but that they omitted the corresponding operation, that is "Display".

		Pre-test					Post-test				
		Hierarchy levels					Hierarchy levels				
	Uninterpretable	1.1	1.2	1.3	1.4	2.1	1.1	1.2	1.3	1.4	2.1
13 year olds	Enumeration of fields	2	3	3	2	3	8	7	5	4	7
	No response and others	0	1	3	5	6	0	0	0	2	2
16 year olds	Enumeration of fields	2	2	2	1	2	4	4	4	2	2
	No response and others	2	0	2	4	3	0	0	0	0	0

Table 8.1 Children's understanding of the nature of data search and structure: Unclassified responses: 13 and 16 year olds: pre- and post-tests

It should be said that there is some consistency in these responses. For example one 13 year old used the "naming fields" strategy in all questions of both pre- and post-test, and four 13 year olds and two 16 year olds did so in all the questions of the post-test.

The fact that, although not as often, this type of response was also apparent in the pre-test may suggest an understanding of the computer not as a processor of information, but as a mere storer of information. In this sense, the computer would not differ much from a book, where information is not found by processing.

8.3.1 How data is stored and linked

8.3.1.1 Younger students

The results from the pre-test show that overall the younger children tend to think that data is just kept somewhere in the computer, and that data is not linked at all or is merely somehow linked, as explained above (see Table 8.2). However, concerning the question about sorting data, eight students seem to understand that data is "record-field" linked in the sense that by sorting on a field (e.g. price) records are re-arranged.

	Hierarchy levels				
	1.1	1.2	1.3	1.4	2.1
How is data stored?					
just kept somewhere	19	17	15	17	15
in an order	5	5	5	2	2
How is data linked?					
not linked	14	15	13	11	8
somehow linked	9	7	6	0	8
record-field linked	1	0	1	8	1

Table 8.2 Children's understanding of how data is stored and linked: 13 year olds: pre-test

Examples to illustrate the most common kinds of responses are as follows:

The following responses to question 1 and 2 respectively would imply that data is seen just as being kept somewhere, and is seen as "not linked":

> *"Go through all the titles until it gets to Jane Eyre"*
>
> *"The computer would search through the fairy tales making a list"*

For question 3 (the retrieval of objects with more than one given characteristic) responses like

> *"Pick out the books in alphabetical order under £5"* , and
>
> *"The computer could search under all the types of books to find all the textbooks, then going through the prices finding all the ones under £5"*

show an understanding respectively, of data as "kept in order" but "not linked", and as "just kept somewhere" but "record-field" linked.

For question 4 (sorting data), here are some examples:

> *"Look through the books prices and put the cheapest ones at the bottom and the expensive ones at the top"* (just kept somewhere - not linked)
>
> *"Go through in alphabetical order to find the most expensive one"* (data is kept in an order and not linked)

Examples of responses for question 5 *("Could the computer find out if the most expensive books usually have the most pages"):*

> *"Look at number of pages and see which ones have the most pages")* (data is seen as just kept somewhere and not linked)
>
> *"Make a list of all the expensive books, then see how many pages they have"* (just kept somewhere and somehow linked)

References to specific locations of the computer were rarely made (they were non-existent in the post-test), and if so, these were as general as "memory" and less commonly "screen". Some examples are:

> *"It would go through its memory until it found all the information about 'Jane Eyre'"*
>
> *"Find the book 'Jane Eyre' in memory"*
>
> *"List the fairy tales section of the memory"*

It should be said that the metaphor used in the questionnaire to describe the organisation of information to be stored in the data base, namely that of the card filing system *("Suppose you have cards with information about children's books . Each card says: Title of the book, kind of book, etc. The information on these cards is put into a data base on a computer .....")* was not very effective, since it seems to have led children to confusion. It was very common to find responses with particular reference to the cards, but only in the pre-test:

> *"It would go through all the cards.........."*
>
> *"Go to the card of "Jane Eyre".........."*
>
> *"By going to the card titled prices.........."*
>
> *"Look at price card and number of pages card.........."*

Results from the post-test show that this time fewer 13 year old students think information is kept in a particular order and not linked, and there is a substantial increase in the number of students who think data is "record-field" linked (see Table 8.2). Such an increase is particularly noticeable for question 1 (one item retrieval) for which more than half of the students thought data was not linked in the pre-test and less than one third thought so in the post-test, and also for question 5 (finding of a relationship) with only one student who saw data as "record-field" linked in the pre-test, but most students doing so in the post-test.

	Hierarchy levels				
	1.1	1.2	1.3	1.4	2.1
How is data stored?					
just kept somewhere	15	17	20	19	16
in an order	2	1	0	0	0
How is data linked?					
not linked	5	11	6	8	2
somehow linked	4	1	9	0	0
record-field linked	8	6	5	11	14

Table 8.3 Children's understanding of how data is stored and linked: 13 year olds: post-test

Examples of responses from the post-test for questions 2 and 3 respectively, which as in the pre-test, show an understanding of data as "just kept somewhere" and "not linked" are given below:

"Go through all fairy tale book files"

"Search for books costing less than £5"

However, in the post-test, a common response, showing an idea that data is kept somewhere and linked, was to display the required characteristics or fields for all the records. For example, for question 2:

"Display records
choose kind of book
title of book
display"

This common type of response made me wonder whether it was relevant to add a new feature within the "how data is stored" feature. Could I infer from a response like that above that children understood data as stored by feature, that is, in columns (or fields). This idea was not in fact used; it is mentioned here to illustrate further problems of the analysis.

Examples from the post-test, regarding question 4, are:

> *"Display records*
> *title of book*
> *price list*
> *cheapest first"* **(data is seen** as stored somewhere in the computer and **"record-field" linked)**
>
> ***"Put the cheapest book you know of into the computer, then see if it knows any others that are any higher"*** (data is seen as not linked)

and regarding question 5:

> *"Display records*
> *title of book*
> *number of pages*
> *price"* (record-field linkage)
>
> *"Display*
> *number of pages"* (not linkage is shown)

8.3.1.2 Older students

The older students do not think data is kept in an alphabetical order as some 13 year olds in the pre-test often do. Most students again tend to think data is "just kept somewhere" (when no particular reference to "where" is made and also when general words like "memory", "data bank", and "cards" are used) in both pre- and post-tests. Data is seen as "not linked" and "somehow linked" in the pre-test, but unlike the younger students, more 16 year olds think data is "somehow linked" than "not linked" at all. Data is seen as mainly "record-field" linked in the post-test (see Table 8.4).

	Pre-test					Post-test				
	Hierarchy levels					Hierarchy levels				
	1.1	1.2	1.3	1.4	2.1	1.1	1.2	1.3	1.4	2.1
How is data stored?										
just kept somewhere	17	23	21	20	20	16	16	16	18	18
in an order	4	0	0	0	0	0	0	0	0	0
How is data linked?										
not linked	8	8	8	10	7	3	8	3	6	2
somehow linked	12	13	11	0	12	5	0	4	0	4
record-field linked	1	2	2	10	2	8	8	9	12	12

Table 8.4 Children's understanding of how data is stored and linked: 16 year olds: pre- and post-tests

The following examples illustrate the general nature of the shift shown in Table 8.4, contrasting typical pre- and post-test answers:

Q1:

pre-) *"Search through the list of book titles in its memory or database until it found "Jane Eyre". It would then have the authors name available"* (data is just kept somewhere and "somehow linked")

post-) *"Display records on titles of books and names of authors"* (data is seen as "record-field" linked)

Q2:

pre-) *"Find the title of book under section of Fairy Tale"*

post-) *"Select kind of book, choose fairy-tales and then display records"* ("record-field" linkage)

Q3:

pre-) *"Ask it to list all the books that are textbooks. List all of them that are under £5"* (data is seen as "somehow linked")

post-) *"It will look up "Kind of Book" and searches for textbooks. Then it looks up "prices" and searches for textbooks less than £5"* (data is seen as "record-field" linked)

Q4:

pre-) *"Find the lowest number and work up in numerical order until all the books have been accounted for"* (data is not linked)

post-) *"Sort*
price
cheapest first
display data" ("record-field" linkage)

Q5:

pre-) *"Give a list of all the books that are expensive and cheap and then list the number of pages they have so we can compare them"* (data is seen as "somehow linked")

post-) *"display records*
price and number of pages
display" (data is seen as "record-field" linked)

8.3.2 How data is found

Students' ideas concerning how the computer can do particular data base operations depend on the requirements of each particular task (or level in the hierarchy). Thus, each question will be analysed separately.

8.3.2.1 The retrieval of a characteristic of an object

Results for question 1: *"How can you imagine the computer could find the name of the author of the book titled "Jane Eyre"* from pre- and post-tests and for the 13 and 16 year olds can be seen in Table 8.5.

	13 year olds		16 year olds	
STRATEGY	pre-test	post-test	pre-test	post-test
• look for target directly	8	2	4	3
• look for known features	16	5	17	6
• show both fields	0	7	0	8

Table 8.5 Children's understanding of how data can be found in a data base for question 1 (level 1.1): 13 and 16 year olds: pre- and post-tests

Results for the 13 and 16 year old students are very similar. The results show that whereas in the pre-test, students (more than two thirds of them) tend to think that the computer would first look for the known features (i.e. "title of the book") and then get the unknown feature associated with it (i.e. "name of the author"), which we might call a "mapping" strategy, in the post-test there is a tendency to think the computer would show the information contained in the two relevant fields (i.e. "title" and "name of the author") on the screen for you to go through. About half of the 13 and 16 year old students in the post-test thought the computer would just display items. To a certain extent, the results show a shift from a "search" strategy in the pre-test, towards a "display" strategy in the post-test. Children seem to have learned through training and practice with GRASS that it is easier to find an item of information by displaying data and then eye-balling, than by doing a search, or maybe they just found the former strategy easier to remember and express on paper.

Examples of the different strategies children attributed to the computer when retrieving a characteristic of an object will be shown subsequently. These will include responses from the 13 and 16 year olds and from both pre- and post-tests.

Look for target directly:

13 pre-) *"It will search through the data banks of authors in alphabetical order"*

13 post-) *"Go through the authors alphabetically"*

16 pre-) *"It will contact the various authors and find out whatever book they wrote"*

16 post-) *"By searching author and book files"*

Look for known features:

13 pre-) *"Find the book "Jane Eyre" in memory. Look up its author"*

13 post-) *"By finding title of book "Jane Eyre" and then showing its author"*

16 pre-) *"Go to the card of "Jane Eyre" and go down the information to find name of the authors"*

16 post-) *"It will look up the field name "title of the book" and searches for "Jane Eyre". It will then look up the author"*

Show one/both fields:

13 post-) *"Display records*
author of book
name of author"

16 post-) *"Choose title of books and authors, press space bar until "Jane Eyre" appears, shows author.*

8.3.2.2 The retrieval of objects with a given characteristic

The search strategies for question 2: *"How can you imagine the computer could find the titles of all the fairy tales"* that children thought the computer would use are shown in Table 8.6.

	13 year olds		16 year olds	
STRATEGY	pre-test	post-test	pre-test	post-test
• look for target directly	9	2	4	2
• look for known features	12	10	19	14
• show one/both fields	1	6	0	0

Table 8.6 Children's understanding of how data can be found in a data base for question 2 (level 1.2): 13 and 16 year olds: pre- and post-tests

As can be seen from Table 8.6, overall most 13 and 16 year olds thought the computer would use a "mapping" strategy, (as previously: the computer returns unknown features in some column which are associated with a known feature in another column) in both pre- and post-tests. However, although more than half of the younger students saw data as being retrieved through "mapping" in the pre-test (the rest thought the computer would look for the target, i.e. titles, directly), one third thought the computer would just display the relevant characteristics for all the records in the post-test. This shift from a "search" strategy to a "display" strategy was not apparent in the responses of the older students. Most of them thought the computer would look for a known feature "fairy tales" within the "kind of book" field in both pre- and post-tests. It seems that children understand that finding several items from a list is not as easy or efficient as just finding one.

Examples of the strategies children imagined the computer would use for the retrieval of objects with a given characteristic are provided next:

<u>Look for target directly:</u>

13 pre-) *"Goes through all the titles and picks out the fairy tales"*

13 post-) *"Go through all the titles and find the kind of book you want"*

16 pre-) *"Look up all the titles of books seeing which ones are and which are not fairy tales and eliminate all the ones which are not"*

16 post-) *"Calling up fairy tale title file"*

<u>Look for known features:</u>

13 pre-) *"Go to kind of book section and search through fairy tales section"*

13 post-) *"Go to list of kind of book, choose fairy tales then ask for the title of the book and list out all the information"*

16 pre-) *"It looks for a kind of book with Fairy-tales. Then it will ask for the title of these books"*

16 post-) *"It will look up the field name "kind of book" and searches for "fairy tales". It will then print out the titles"*

<u>Show one/both fields:</u>

13 pre-) *"It will show you what each kind of book was (fairy tales, etc.)*

13 post-) *"Display records*
title of book
kind of book"

8.3.2.3 The retrieval of objects with more than one given characteristic

Results for question 3: *"How do you imagine the computer could find all the textbooks which are cheaper than £5"* are shown in Table 8.7.

STRATEGY	13 year olds		16 year olds	
	pre-test	post-test	pre-test	post-test
• look for target directly	5	0	0	0
• look for one known feature	12	8	18	10
• look for two known features - separately	3	6	3	2
- combine them logically	0	2	0	2
• show one/both fields	0	4	0	2

Table 8.7 Children's understanding of how data can be found in a data base for question 3 (level 1.3): 13 and 16 year olds: pre- and post-tests

Table 8.7 shows that most 13 and 16 year old students said the computer would focus just on one known feature (e.g. price less than £5), instead of on both. This may be due to difficulties in dealing with two variables at a time, or it may also be due to difficulties in understanding the question. In Question 3 the word "kind of book" was missing, and perhaps children understood "textbooks" not as a kind of book but as all the books in the file. It should be noticed, however, that this problem did not exist for question 2 (the finding of all the fairy tales), in which "kind of book" was also not present. Another reason could be that children are not simply aware of the capabilities of a data base, and cannot even think the computer can do complex searches. Assuming that it can only deal with the retrieval of objects with one characteristic, they took "cheaper than £5" (which may be more explicitly expressed than "textbooks") as the condition.

Just a few students said the computer would look at the two variables, and when so, they mostly said it would do it separately. Just two 13 year olds and two 16 year olds, and only in the post-test, thought the computer would combine the two variables logically.

The "show" or "display" strategy was not used in the pre-test by neither 13 nor 16 year olds. However, it was used in the post-test, but by very few students. For complex searches, the "show" strategy (displaying the two relevant characteristics for all the countries) is used less than it is for simple searches (see Table 8.5).

Examples of the search strategies shown in children's responses to question 3 can be seen below:

Look for target directly:

13 pre-) *"Go through alphabetically until you find them"*

Look for one known feature:

13 pre-) *"Go to price section and search through finding books under £5"*

13 post-) *"Search book costing less than £5"*

16 pre-) *"Find a list of prices of all the books, then tell it to list all those less than £5"*

16 post-) *"Search data, book less than £5 display"*

Look for two known features separately:

13 pre-) *"Find all the textbooks, go through prices to see all the ones under £5"*

13 post-) *"Go to kind of book, choose textbook, then go to price and list out the information"*

16 pre-) *"Ask the computer to go to the kind of book and search through all the cards and find textbooks, then go through price and display"*

16 post-) *"Sort data on prices, choosing smallest first then display records on textbooks"*

Combine the two known features logically:

13 post-) *"Search data*
price
less than £5
narrow down the search
kind of book"

16 post-) *"Search data*
kind of book
textbook
narrow down search to price of book less than £5
print name of book, author and price"

Show one/both fields:

13 post-) *"Display records*
title of book
kind of book
less than £5"

16 post-) *"Sort data*
kind of book
display records
kind of book
price"

8.3.2.4 Sorting data on a given characteristic

Most students seem to take it for granted that the computer can sort data. Their responses do not often include explanations of what the sorting activity is about, but show an understanding of "sorting" as an implicit operation of the system ("the computer will just do it"). The "extreme" strategy (the computer looks for the highest and/or lowest value of a characteristic) is more commonly used by the 13 than by the 16 year olds, and mainly in the pre-test. Table 8.8 shows these results.

	13 year olds		16 year olds	
STRATEGY	pre-test	post-test	pre-test	post-test
• show field	3	3	5	0
• "extreme"	7	3	1	4
• sort	9	13	14	14

Table 8.8 Children's understanding of how data can be found in a data base for question 4 (level 1.4): 13 and 16 year olds: pre- and post-tests

Examples:

Show one/both fields:

13 pre-) *"By looking at the books name list and at price list, and then showing out"*

13 post-) *"Display records*
title of book
price list
cheapest first"

16 pre-) *"By going to the card titled prices then displaying them on the screen"*

"Extreme":

13 pre-) *"The computer could find the cheapest book and the most expensive book, then finding the books nearest to each book making two lists, then add one to the other"*

13 post-) *"Put the cheapest book you know of into the computer, then see if it knows any others that are any higher"*

16 pre-) *"Find the lowest number and work up in numerical order until all the books have been accounted for"*

16 post-) *"Find the cheapest and lowest price*
list in order of the range"

Sort:

13 pre-) *"Look at "price" and programme it to do put them in order"*

13 post-) *"Sort prices*
cheapest first
display titles"

16 pre-) *"It will look up Price. Then it will sort through all the books in the whole file and place them in order from cheapest to most expensive"*

16 post-) *"Sort data*
price
cheapest first
display records
price
title of book"

8.3.2.5 Finding a relationship between two variables

Results for the 13 and 16 year olds show a very similar pattern. In the pre-test, children tend to use the "extreme" strategy, and in the post test a "(sort+) show" strategy and a "graph" strategy were equally common. These results are shown in Table 8.9.

STRATEGY	13 year olds		16 year olds	
	pre-test	post-test	pre-test	post-test
• "extreme"	14	4	12	3
• (sort +) show	3	6	6	8
• graph	0	6	2	7

Table 8.9 Children's understanding of how data can be found in a data base for question 5 (level 2.1): 13 and 16 year olds: pre- and post-tests

Perhaps, the "(sort +) show" strategy should have been divided into two: a "show" strategy and a "sort + show" strategy, since the latter is much more efficient than the former. It should be said that none of the 13 year olds thought the system would be used to organise the data ("sort + show" strategy) in the pre-test, and two (out of six) did in the post-test. The 16 year olds were better at thinking the computer would sort the data before displaying it on the screen. Three did it in the pre-test and five in the post-test.

Out of the fourteen 13 year olds who thought the computer would use an "extreme" strategy in the pre-test, three said it would just look at one extreme of one variable, as can be observed in the following example:

> *"Look at number of pages and see which ones have the most pages"*

Another three said the computer would look at both extremes and at both variables:

> *"It would list expensive books, then show how many pages each book has. Then show cheaper books and their pages"*

And the rest, eight students, thought the computer would focus on just one extreme and on both variables:

> *"Make a list of all the expensive books, then see how many pages they have"*

None of the 16 year olds, however, said the computer would look at just one variable, and most (nine out of the twelve who used the "extreme" strategy) said it would look at the higher values of one variable (usually price) and then examine the relation between that variable and the second variable (i.e. number of pages) by displaying both variables on the screen. This was expected, since the question itself *("If the most expensive books usually have the most pages")* focuses on the higher values. Perhaps, if children had been asked for a relation between price and number of pages, they would have focused on the lower values too.

The "extreme" strategy, although efficient, is not as efficient as the "(sort +) show" strategy or "graph" strategy, which demand less interaction with the system, but it is more efficient than the simple "show" strategy. Overall, children's conceptions about how the computer could find relationships between two variables became more useful after having learned GRASS (post-test). Children learned that the computer could use two strategies ("graph" and "sort + show") which are efficient and require little time to be completed.

Some examples of the strategies described in the students' responses are shown below:

<u>"Extreme":</u>

13 pre-) *"Find books with the most pages"*

13 post-) *"List expensive books. Give the pages each book has"*

16 pre-) *"By searching for the biggest books and comparing them to price"*

16 post-) *"It will look up "price" and then find books over a certain given price. Then it will look at the "number of pages" and print out the price and pages. Then it will look up "price" and searches for books under a given price, then the number of pages and print them out. Now they can be compared"*

<u>(Sort +) show:</u>

13 pre-) *"List number of pages in each book"*

13 post-) *"You'd sort data and have expensive books first. Then you'd ask it to display number of pages and price"*

16 pre-) *"List all books from cheapest to the most expensive including and highlighting number of pages and type of binding"*

16 post-) *"Sort data on prices, most expensive first. Then display records on prices and number of pages"*

<u>Graph:</u>

13 post-) *"Graphs and figures*
scatter diagram
price and pages"

16 pre-) *"Look at price card and number of pages card, so it can show the relationship in a table or something"*

16 post-) *"Graph*
scatter graph
vertical axis price
horizontal axis no. of pages
display"

8.3.3 What type of language is used?

All 13 and 16 year olds used ordinary language in the pre-test, and about half of them used the GRASS query language, although not always successfully, in the post-test. As an exception, almost all 13 year olds (fourteen out of sixteen) used a data base query language in the post-test for the question involving relationship (see Table 8.10). Children may have learned by using GRASS that the computer can only find information if you tell it how to do it, and in a language it can understand.

		Pre-test					**Post-test**				
		Hierarchy levels					Hierarchy levels				
	Type of language	1.1	1.2	1.3	1.4	2.1	1.1	1.2	1.3	1.4	2.1
13 year olds	Ordinary language	24	22	20	21	21	9	10	8	7	2
	DB query language	0	0	0	0	0	8	8	12	13	14
16 year olds	Ordinary language	21	23	21	20	20	7	7	7	9	8
	DB query language	0	0	0	0	0	9	9	9	9	10

Table 8.10 Type of language used: 13 and 16 year olds: pre-and post-tests

The vocabulary of GRASS (Display, Search, Sort....) and the grammar or format rules (e.g. SORT is followed by the name of a field and an option: biggest first or smallest first) was acquired with relative success by about half of the students. Analysis of the errors in the queries children formulated shows their main difficulty in distinguishing between the operations "sort" and "search", and "search" and display". Here are some examples:

> *"Sort data*
> *less than £5*
> *display data*
> *book*
> *price"*

> *"Display*
> *kind of book*
> *is the same as fairy tale*
> *display"*

The question of whether understanding of the syntax of GRASS really means understanding of the nature of data bases in general is not clearly answered. From a conceptual point of view, responses of the kind:

> *"Go to titles of books showing authors with it"*

> *"Look for card with price of books, then sort them running from cheapest to most expensive"*

from the pre-test, are respectively similar to the following responses from the post-test:

> *"Display records*
> *title of book*
> *name of authors"*

> *"Sort prices*
> *cheapest to expensive*
> *display information"*

8.3.4 Does method work?

The effectiveness of the different search strategies children imagined the computer would use for solving the various data base tasks has already been discussed in section 8.3.2. Also, brief reference to the most common types of error children made when writing their data base queries in the GRASS syntax has been made in section 8.3.3.

8.4 Forming questions

At the end of the questionnaire, children were asked to suggest questions about the books they thought the computer could answer. Many students did not answer this last question (nine 13 year olds in the pre-test and eleven in the post-test; fifteen 16 year olds in the pre-test and ten in the post-test), but others wrote more than one question. Overall, the results show that almost none wrote questions concerning the retrieval of objects with more than one given characteristic (level 1.3 in the hierarchy), and that whereas 13 year olds tend to write questions involving sorting data and finding relationships (levels 1.4 and 2.1 respectively), 16 year olds focused mainly on the retrieval of objects with a given characteristic (level 1.2) (See Table 8.11). Surprisingly, "uninterpretable" questions, such as follows, were very common even in the post-test:

> *"Books which are most popular to a different age group"*
>
> *"Summary of the events in the book"*
>
> *"Number of words per book"*
>
> *"Authors in age order because it knows all about the author"*

Within this category, the simple enumeration of characteristics was also existent:

> *"Type of binding*
> *date of publishing*
> *publishing firm*
> *number of pages"*
>
> *"Number of pages*
> *subject*
> *author*
> *availability"*

Hierarchy levels	13 year olds		16 year olds	
	pre-test	post-test	pre-test	post-test
1.1	4	2	1	1
1.2	4	1	4	6
1.3	1	1	0	0
1.4	6	3	1	1
2.1	5	4	3	3
Uninterpretable	7	4	6	8

Table 8.11 Number of questions under each hierarchy level children asked of the data base

These results show the difficulty children have in forming meaningful questions for a data base despite a reasonable understanding of the nature of data and of data base capabilities in general. Some children seem not to understand that the computer can only find data which has previously been stored.

Examples of questions children asked are shown below:

level 1.1

13 pre-) *"How many chapters are there in "Noddy met Big Ears"?"*

13 post-) *"How many pages does the book by........ have?"*

16 pre-) *"Number of illustrations in the book 'Jane Eyre'"*

16 post-) *"Price of the book by a certain author"*

level 1.2

13 pre-) *"Find Enid Blyton's books"*

13 post-) *"What are the titles of books under £10?"*

16 pre-) *"Given the author's name, find the titles of books under his name"*

16 post-) *"Books published after certain date"*

level 1.3

13 pre-) *"How many fairy tales are in hardback edition"*

13 post-) *"List all the short-stories of a price under £2"*

level 1.4

13 pre-) *"What is the oldest book?"*

13 post-) *"What is the largest size?"*

16 pre-) *"To put the books in order of size"*

level 2.1

13 pre-) *"Does one particular publishing firm specialise in types of books?"*

13 post-) *"Does the price increase according to the number of illustrations?"*

16 pre-) *"Does the publishing firm affect price?"*

16 post-) *"Is the number of pages larger if there are a lot of illustrations?"*

8.5 Some conclusions

Before learning and working with a data base, children hold an understanding of data as just being "kept somewhere" in the computer, and not linked by means of records and fields. For retrieval tasks, both 13 and 16 year olds think data can be retrieved from a data base by looking for the known feature/s and getting the unknown features associated with them (simple mapping). One third of the younger students, however, tend to think the computer would simply look for the unknown feature directly in the task requiring the retrieval of an item. For questions concerning the retrieval of objects with more than one given characteristic, students do not think the computer would look for the target directly, but it would look for just one of the known features, not paying attention to the other characteristic. Most children think the computer can sort data by simply doing it. A few 13 year olds, however, thought the computer would first find the highest and/or lowest value of a characteristic ("extreme" strategy). Both 13 and 16 year olds think that relationships between two variables can be found by looking at the higher values of one variable and then examining the relation between that variable and the second one.

After having learned how to use GRASS and having solved a number of data base tasks, children's ideas about how searches are done in the computer have changed. First, regarding one-fact retrieval tasks, children change from a "mapping" strategy to tending to think that the computer would just display the information of the two relevant characteristics on the screen, so that you can go through it and find the item you need. This new strategy is, in fact, less efficient than the previous one. When the task requires the retrieval of more than one item, children think that, as in the pre-test, the computer would first look for the known feature and then return the unknown values associated with it. A few 13 year olds, however, changed to saying that the computer would use the less efficient strategy of displaying both variables on the screen for you to go through. For tasks requiring the use of logical operators AND/OR, both in pre- and post-tests children think the computer would just look up one characteristic, but a few more 13 year olds changed to saying the computer would search for the two variables, although separately. Sorting is seen as possible by using a "Sort" command followed by the name of the characteristic to sort on. Regarding the finding of a relationship between two variables, whereas most 13 and 16 year olds in the pre-test thought the computer would use an "extreme" strategy, most students changed to thinking that the computer would either display a scattergraph of both variables on the screen ("graph" strategy) or display both variables simultaneously on the screen ("show" strategy) or, sort data on a variable and then display the two ("sort + show" strategy).

Results from the post-test show that children still think data is "just kept somewhere" in the computer (at least, no information about how data is stored is provided in their responses), but that some have changed to thinking that data is "record-field" linked. The syntax of GRASS (vocabulary and grammar) was only used in the post-test, by about half. It is not clear whether this really means that their general understanding of the nature of data bases improved.

Chapter 9 CONCLUSIONS

9.1 Introduction

This research hopes to have thrown some light on the issue of children's learning and use of computer data bases. It has attempted to contribute by presenting some empirical data on how well children can learn and use data bases, how well they understand the nature of data bases, and how well they grasp the relevant logical operations. This final chapter is an attempt to summarize the findings of the research and to present the main conclusions which derive from them.

9.2 Overview of the research

The research focused on two important elements of the teaching and learning of data bases: the nature and structure of data base tasks, and the skills required for them.

For the analysis of the tasks, a "logical" hierarchy of data base tasks was defined (see section 2.6). The hierarchy distinguished two types of tasks and six sub-tasks, as follows:

1 Tasks which require the retrieval of specific data

- 1.1 The retrieval of a characteristic of an object
- 1.2 The retrieval of objects with a given characteristic
- 1.3 The retrieval of objects with more than one given characteristic
- 1.4 Sorting data on a given characteristic

2. Tasks which require reasoning on the data retrieved

- 2.1 Finding a relationship between two variables
- 2.2 Finding relationships between all the variables

This "logical" hierarchy of tasks was tested with students of 13 and 16 years of age, after having taught them how to use a simple data base. Two different age groups were chosen to see whether the hierarchy was apparent in their work, and if so, how it may depend on age.

The data base used in the study was GRASS, a menu-driven program. GRASS was chosen because it is currently one of the most commonly available data bases in English schools, and is one of the few educational data bases that allow graphical representations. Another reason for the choice was its simplicity and ease of use.

The tasks used to test the "logical" hierarchy of tasks consisted of two sets of five data retrieval tasks posed in form of questions (one for each level of the hierarchy) (see section 3.2.1.1), and a "report editing" task. For the latter, the five different levels of tasks were incorporated in form of ten statements embedded within a report for children to check and evaluate (3.2.1.2). In all, each child solved four tasks at each hypothetical level of difficulty. It was expected that children would do better in those tasks which required the retrieval of specific data than on those which required reasoning on the data retrieved.

Regarding the skills and knowledge required for solving the tasks that the children possessed, two paper and pencil tasks were designed to test children's understanding of logical operators (see section 3.2.2):

- An interpreting task, requiring the interpretation of a logical sentence as a pattern of search
- A constructing task, requiring the building up of a logical sentence to achieve a desired pattern of search

Two other questionnaires were designed to test children's understanding of the nature of data bases. One concerns children's understanding of the types of information that can be retrieved from a data base, and the second concerns their understanding of how the computer works internally, how it stores and finds information (see sections 3.2.3 and 3.2.4). Children were tested before and after having been taught about GRASS and having solved the data base tasks (the two sets of exercises and the "report-editing" task) interacting with the computer.

9.3 Data base tasks - Empirical results

9.3.1 "Empirical" hierarchy of data base tasks

The analysis of children's performance on the data base tasks used in this research supports some hierarchy of tasks, but not in the detail theoretically proposed. The "empirical" hierarchy of tasks suggests the existence of two levels of tasks:

(i) *easy tasks* (all being about equally easy), accomplished by the great majority of children at both ages:

- the retrieval of a characteristic of an object
- the retrieval of objects with a given characteristic
- sorting data on a given characteristic

(ii) *harder tasks* (those involving logic being slightly harder than those involving relations), which show some improvement with age, and require dealing with more than one variable at a time:

- the retrieval of objects with more than one given characteristic
- the finding of a relationship between two variables

The "empirical" hierarchy suggests that children can learn to use a simple data base easily and with almost no teaching, and can perform a number of simple data base operations successfully, even at a relative young age. They may have difficulties, however, with those tasks known to be relatively difficult, such as the two harder tasks of the hierarchy (logic and relations). Common errors such as attention paid to just one condition or variable, or to both but separately, and the selection of the wrong fields to search may be expected when children are given these tasks.

Selection of the wrong fields to search was particularly common for the younger students. These results appear in chapter 5.

Children's difficulties with tasks involving logic were expected. The interpretation of logical sentences was shown to be easy for AND sentences but hard for negative OR sentences (see section 6.1.1).The construction of logical sentences was shown to be much harder than interpreting them (see section 6.1.2), especially for the younger students. The logical operators AND and OR are not always understood in their logical sense. Also, children's particular difficulty in suggesting questions involving logical operators (see section 8.4), and children's model of the search system as able to focus on one only variable when retrieving objects with more than one given characteristic (see section 8.3.2.3), both help to account for children's relatively poor performance in this type of tasks. The testing of relationships also proved to be difficult for the younger group. Results from the questionnaire on children's understanding of the nature of data reveal that the older group understood data bases as able to deal with the finding of relationships between two variables substantially better than the younger group (see section 7.1.2).

9.3.2 "Real life" tasks (see EGGLESTON, 1983)

Although research (e.g. APU) shows that "real life" tasks are often harder than more artificial tasks (tasks in the form of structured exercises), which are usually more direct, the analysis of the "report editing" task showed that data base tasks embedded within real life contextualised tasks need not be particularly hard. Problems with language existed, as expected, but the "report editing" task could be done well by children of both ages. These tasks, developed for the research, could well be of pedagogic value to teachers .

9.3.3 Types of search strategies used

There appear to be two common different information retrieval strategies, but students were not consistent in their preference for one or the other. The most efficient strategy which retrieves just the information required, and no more, tended to be used for the tasks which required the retrieval of objects with a given characteristic, the retrieval of objects with more than one given characteristic, and sorting data on a given characteristic. The less efficient "show and find for yourself" strategy, where the child does some of the work of retrieval, tended to be used for retrieval of a single fact. Finding an item from a list seems to be easier than finding several. Both strategies seem to be equally employed for the finding of relationships between two variables (see sections 5.2.1, 5.2.2 and 5.3.2).

The results indicate that the computer is not always used to its full advantage and that there is a tendency to use the computer to do part of the work, leaving some work for the child to do. This requires the use of simpler commands (DISPLAY instead of SEARCH) and only a simple understanding of the computer as able to store data and show it in the form of a list, leaving to the user the task of checking and counting, and not requiring an understanding of the computer as able to do logical calculations. These findings are confirmed by results of the post-test on children's understanding of the nature of data search and structure (see section 8.3.2). Children's previous ideas of the nature of data search, however, are more concerned with the computer as a "search" system than as a "display" system.

9.4 Data base use - A developmental issue?

A comparison of performance of the two age groups on the data base tasks shows that there is no fundamental difference between them except that the older group does a little better in the harder tasks, particularly in the "report editing" task, which requires children to think about and find questions that could be asked about the text.

The results suggest that there is a relative improvement in performance on the harder tasks after three years, but since most students of both ages can do a number of simple data base tasks (easy tasks) successfully, and about half of the younger students can also do the harder ones, it can not be asserted that a simple developmental progression regarding data base use exists as such. The fact that the older group outperform the younger group in the "report editing" task (see section 5.4) and did not have so many problems with the selection of the fields to search (see section 5.2.2) may be due to their better command of language. Also, results from the Constructing task show that the older students can construct negative AND sentences better than the younger ones (see section 6.1.2.2). Other than in this aspect, older children did no better in logical tasks than the younger ones, suggesting that if there is a developmental progression, it is far from complete at age 16. The fact that the older students could decide about the existence or not of a relationship, after having displayed a scattergraph, better than the younger students also suggests their better understanding of graphical representations (see section 5.2.2).

9.5 Logic, language, and data base use

The analysis of the Interpreting and Constructing tasks shows that children have difficulty in interpreting and constructing logical sentences, particularly OR sentences, contradiction and tautology. The task of building up a structure of data, given a logical sentence, seems to be easier than the task of building up a logical sentence given a particular structure of data. One explanation of this result could be that it is linguistic in origin, supported by results indicating the use of inclusive OR as if it were exclusive OR, and the use of OR as AND when negation is involved. In this case, AND almost completely replaces OR, perhaps used in the sense "AND AS WELL" (AND as logical addition, that is to say OR). It seems that children's understanding of the meaning of logical connectives in natural language (conjunctions are the most frequently occurring connectives; disjunctions are much more rarely formulated and heard by children) influences performance on data base tasks. English language knowledge seems to confuse instead of to facilitate data base tasks. Another explanation could be the need in the sentence construction task to see the pattern of filled and empty cells as a whole, and then to convert it into a logical sentence, by contrast with the interpretation task in which cells can be ticked one at a time. Children's visual interpretation of the search patterns seems to affect the construction of logical sentences, particularly the tautology (all cells are filled), which is associated with affirmative AND logical sentences, and contradiction (empty cells), which is associated with double negative AND sentences (see section 6.1.2). A third interpretation could be related to children's greater difficulty in interpreting logical patterns than logical sentences. Maybe children have particular difficulties in relating intuitive visual patterns to verbal deductive forms. These explanations need more attention in further research.

The influence of language in data base use is also reflected in the analysis of the two sets of exercises, and of the "report editing" task. It was especially notable in the last, where children had to deal not only with the analysis of sentences, but with the analysis of a text. Analysis of the exercises shows that even when solving tasks of the same "logical" level of difficulty, performance is affected by the particular question asked, and by how the question is expressed (e.g. when the names of the fields are made explicit in the sentence, performance is better) (see section 5.2.1). The "empirical" hierarchy of data base tasks shows difficulties in distinguishing more levels of tasks, not only because scores on the three easy tasks were very similar but also because they were sensitive to the specific nature of questions (see section 5.4). Analysis of the "report editing" task shows children's difficulty in making links between sentence meanings, and the treatment of anaphoric clauses as "given", not "new" (see section 5.3.1). Understanding a text also involves a good deal of inferencing, observed when children call upon their knowledge of the world to answer particular statements (see section 5.3.1).

Thus, children's general performance on data base tasks may be influenced by the nature of the task. What may be easy in one task, may become difficult in another, despite their making the same logical and other formal demands. An example is an "easy" task - the retrieval of objects with a given characteristic - which became very difficult in one case in the "report editing" task, when the statement to be checked was implied grammatically rather than given explicitly. Children's understanding of the types of questions that can and can not be answered by a data base was also sometimes dependent on the content of the questions asked, again even when the questions were of the same "logical" level of difficulty (see sections 7.1.1 and 7.1.2).

Analysis of the tasks also suggests a possible difficulty children may have in verbalising those data base operations which involve the use of logic and the testing of relationships, even when they can perform them (see section 5.2.1), suggesting a difficulty with language. The Spearman correlation of hierarchies ("empirical" and "logical") also proved to be higher for success in a task than for the descriptions of the strategy used. Operations which may be easy to perform may not necessarily be easy to describe (see section 5.4).

9.6 Children's understanding of data bases

Children with no experience of data base use seem to have an adequate if simple model of a data base in terms of the types of information a data base can and cannot contain, with the older students being slightly better than the younger ones. Both age groups generally correctly identify information which can be stored or can be obtained from a data base (e.g. a fact or a relationship), although there is some difference between simple facts and relationships, relationships being less well recognised as able to be retrieved (see sections 7.1.1 and 7.1.2). Fewer can correctly identify information which can not be kept in a data base, such as information about Judgment, Future, Context, Feelings or Intention. Performance here varies considerably, information about Future and Intention being better identified as not able to be kept in a data base than information about Judgment, Feelings and Context.

Concerning the types of information amenable to data base analysis, children are clear that a data base can deal with questions involving sorting data and the retrieval of one or several items of information, but they are not so sure about its capability to test relationships, particularly the younger students. These results are consistent with the "empirical" hierarchy of data base tasks.

Children's models of the ways information is kept, organised and found in a data base are not so adequate. Children tend to think that data is "just kept somewhere" in the computer, and not linked by means of records and fields. They mostly think the computer can retrieve data through "mapping" (by looking for the known value and then getting the unknown value associated with it), although this "mapping" is not always treated as perfect, and often there is work left for the user. Sorting is however seen as a possible and inherent operation of the computer. Children understand the computer as able to test relationships by means of a "extreme" search strategy. These mental models for search strategies were not applied consistently when solving the data base tasks (see chapter 5). However, the mental models some children formed of the search system after having learned and practised with GRASS were clear manifestations of their new knowledge of the system. The fact that after the teaching, the computer was understood as able to test relationships by using graphs and figures (scattergraph), is one example.

9.7 Learning a data base and learning about data bases

Results of the present research show that a simple menu-driven data base, such as GRASS, can be learnt in a very short period of time (about an hour) by both 13 and 16 year olds so as to allow them to solve a number of simple data base tasks very successfully, and a number of more complex data base tasks (logic and relationships) moderately well. The teaching, consisting of a simple description of the main characteristics of GRASS and of the main options of the interrogation menu through the presentation of examples in form of questions (see Appendix 3), proved to be enough for the learning of simple operations. Logic and relationships, however, may require much more teaching effort, and (especially in the case of logic) may be very difficult to achieve.

Children did not learn about the nature of the logical operators through GRASS, even though GRASS gives an interpretation of the logical operators AND and OR (narrow or add to the search). GRASS did not help them to better interpret and construct logical sentences. As in the pre-test essentially nearly all students interpreted AND items very well, but OR items, contradiction and tautology very badly in the post-test (see section 6.1.3.1). Some younger students improved in the construction of double negative AND sentences. However, experience with GRASS seems to have helped both groups to better understand the types of questions that a data base can answer (questions involving sorting, retrieval and relationships). It also helped the older students to better understand those types of questions which a data base can not answer. Children also learned to better distinguish between the data base and the metaphor used to explain it. Whereas some children tend to refer to the card filing system when expressing their prior understanding of the nature of data search and structure, none did so after having learned the use of GRASS. Children also learned that the computer can display particular fields for all or part of the records, and that it may be easier to find an item of information by displaying data and then eye-balling, than by doing a search. Only a few learned that the computer could combine conditions logically. More learned, however, that relationships can be tested graphically and also by sorting data on a characteristic and displaying both on the screen, both of which are efficient strategies. About half the students of both ages spontaneously used the GRASS query language to describe searches, although not always successfully. DISPLAY and SORT were more commonly used than SEARCH, perhaps being easier to remember and express on paper. The fact that about half the students used the syntax of GRASS in the post-test, even when it was not requested, suggests an understanding of the computer as able to find information only when you tell it how to do it, and in a language it can understand.

9.8 What advice can we give to teachers ?

First, we would say that it is entirely appropriate to use simple data bases with young children. They learn easily to do simple retrieval operations and to sort data, but they may have problems regarding relationships and logical relations. However, some improvement may be expected in these two areas, but not immediately. The results of this research show some improvement with older children.

Second, data bases provide a meaningful environment to learn about logic and relations, both known to be difficult. Logic should not be taught in terms of AND and OR (the linguistic meaning of these operators can interfere with their logical meaning), but other terms such as containing - not containing, widening - narrowing, or using representations like Venn diagrams. For teaching relationships, more emphasis should be put on the meaning of sorting before displaying and also on understanding displayed data and graphical representations.

This research has contributed by presenting a whole range of data base tasks and instruments, which could be used as interesting teaching instruments. Teachers should not be afraid of using "real life" tasks, such as the "report editing" task. The results show that this can be perfectly viable, and can be done by students of different ages. My own observation is that they also enjoyed the "report editing" task much more than the two sets of data retrieval tasks. The range of tasks here provided could also be used for diagnostic purposes, in different areas of the curriculum and for different age groups.

9.9 What advice can we give to curriculum planners ?

The present research has provided further insight on what children of two different age groups should be expected to know, understand and to be able to do when using a particular data base. Although no difference has been found between the two groups regarding simple operations, some differences regarding more complex operations seem to exist. The results of the research could help curriculum planners to set more specific attainment targets regarding data base use. The common *"select and interrogate a computer data base to obtain information needed for a task"* or *"retrieve information from a data base"* could be more specified and differentiated for each key stage to facilitate assessment. It would be interesting to repeat this study with primary children in order to see if the hierarchy of tasks is apparent in their work in the same way, or whether more levels of tasks appear to be distinguished.

BIBLIOGRAPHY

AUSUBEL, D. P. et al (1968), *Educational psychology: A cognitive view*, London: Holt, Rinehart and Winston.

BEEK, J. V., BEEN, P. & HURTS, K. (1989), Computer-Based Learning For On-Line Data Base Search, *Computers and Education*, Vol. 13, N. 4, pp. 327-336.

BERRY, D. C. (1983), Metacognitive experience and transfer of logical reasoning, *Quarterly Journal of Experimental Psychology*, Vol.35A, pp. 39-49.

BEYER, B. (1984), Improving Thinking Skills - Defining the Problem, *Phi Delta Kappan*, 65, (7), pp. 486-490.

BEYER, B. K. (1988), Developing a Scope and Sequence for Thinking Skills Instruction, *Educational Leadership*, Vol. 45, N. 7, pp. 26-30.

BEZANILLA, M. J. & OGBORN, J. (1992), Logical Sentences and Searches, *Journal of Computer Assisted Learning*, 8, pp. 37-48.

BLEASE, D. & COHEN, L. (1990), *Coping with computers: An ethnographic study in primary classroom*, London: Paul Chapman.

BLOOM, B. S. (1956), *Taxonomy of Educational Objectives. Cognitive Domain*, New York: David Mc KAY.

BOWER, G. (1977), *The psychology of learning and motivation*, (Vol.II), New York: Academic Press.

BOYD-BARRET, O. (1990), School Computing Policy as State - Directed Innovation, in BOYD-BARRET, O. & SCANLON, E. (Eds.), *Computers and Learning*, Wokingham: Addison Wesley, pp. 3-22.

BRAINE, M D. S. & RUMAIN, B. (1983), Logical Reasoning, in MUSSEN,P. H., Handbook of Child Psychology III (Cognitive Development), New York: John Wiley & Sons, pp. 263-340.

BRAINE, M.D.S. & RUMAIN, B. (1981), Development of Comprehension of 'OR': Evidence for a sequence of competences, *Journal of Experimental Psychology*, Vol. 31, pp. 46-70.

BRAINE, M.D.S. (1978), On the Relation Between the Natural Logic of Reasoning and Standard Logic, *Psychological Review*, Vol.85, N.1, pp. 1-21.

BRAZIER, F. M. T. & BEISHUIZEN, J. J. (1988), Can an Interface influence information retrieval?, *Paper presented at the AERA Conference.*, New Orleans.

BRAZIER, F. M. T. & BEISUIZEN, J. J. (1988), The influence of the interface on information retrieval, in COLLINS et al (eds.), *Fifth international conference on technology and education,* Edinburgh: CEP consultants, Vol. 2, pp. 246-249.

BRAZIER, F. M. T. (1991), *Design and evaluation of a user interface for information retrieval*, Amsterdam: Vrije Universiteit.

BRIGHT, G. W. (1989), Data Bases in the Teaching of Elementary School Mathematics, *Arithmetic Teacher*, Vol. 37, N. 1, pp. 38-42.

BROSEY, M. & SHNEIDERMAN, B. (1978), Two experimental comparisons of relational and hierarchical database models, *International Journal of Man-Machine Studies*, Vol. 10, pp. 625-637.

BURNARD, L. (1987), Principles of Database Design, in RAHTZ, S., *Information Technology in the Humanities: tools, techniques and applications*, Chinchester: Ellis Horwood, pp. 54-68.

CAMERON C. T. et al (1988), Electronic Spreadsheets and Data Base Management, in Behrmann, M. M., *Integrating coomputers into the curriculum: A handbook for special educators*, Boston, MA: College-Hill Press, pp. 205-230.

CHEN, M. & PAISLEY, W. (1985), *Children and Microcomputers. Research on the Newest Medium*, CA: SAGE.

CHILD, D. (1978), *The essentials of Factor Analysis*, London: Billing and Sons.

CORNBLETH, C. & FORTH, W. (1981), If remembering, Understanding and Reasoning are important....., *Social Education*, April, pp. 276-279.

CROFT, N. (1990), *An action research based analysis to ascertain the extent to which children can become more autonomous users of data bases and micro computers*, MA thesis, University of London.

DATE, C. J. (1990), *An Introduction to Database Systems*, Vol. 1, Fifth Edition, New York: Addison-Wesley, pp. 3-102.

DEGL' INNOCENTI, R. & FERRARIS, M. (1988), Database as a tool for promoting research activities in the classroom: an example in teaching humanities, *Computers and Education*, Vol. 12, N. 1, pp. 157-162.

DES (1989), *National Curriculum; From policy to practice*, London: Department of Education and Science.

DILLON, J. et al (1989), A national environmental data base in action, *IT and learning*, December, pp. 49-51.

DONALDSON, M. (1959), Positive and Negative information in matching problems, *British Journal of Psychology*, Vol. 50, pp. 253-262.

DURDING, B. M., BECKER, C. A. & GOULD, J. D. (1977), Data Organizations, *Human Factors*, Vol. 19, N. 1, pp. 1-14.

EGGLESTON, S. G. (1983), *Learning Mathematics; How the work of the Assessment of the Performance Unit can help teachers*, London: Department of Education and Science.

ELDER, C. L. & WHITE, C. S. (1989), A World Geography Database Project: Meeting Thinking Skills Head On, *The Computing Teacher*, Vol. 17, N. 3, pp. 29-32.

EVANS, J. St. B. T. (1972), On the problems of interpreting reasoning data: logical and psychological approaches, *Cognition*, Vol. 1, N. 4, pp. 373-384.

EVANS, J. St. B. T. (1972), Reasoning with negatives, *British Journal of Psychology*, Vol. 63, pp. 213-219.

EVANS, J. St. B. T. (1982), *The psychology of deductive reasoning*, London: Routledge & Kegan Paul, pp. 113-208.

FILLENBAUM, S. (1977), Mind your p's and q's: The role of content and context in some uses of and, or and if, in BOWER, G. ,*The psychology of learning and motivation*, (Vol.II), New York: Academic Press. pp. 42-97.

FITTER, M. (1979), Towards more 'natural' interactive systems, *International Journal of Man-Machine Studies*, Vol. 11, pp. 339-350.

FRAAS J. W. (1983), *Basic concepts in educational research*, London: University Press of America

FREEMAN, D. & LEVETT, J. (1986), Quest - two curriculum projects: perspectives, practice and evidence, *Computers in Education.*, Vol. 10, N. 1, pp. 55-59.

FREEMAN, D. & TAGG, W. (1985), Databases in the Classroom, *Journal of Computer Assisted Learning*, Vol. 1, N. 1, pp. 2-11.

GRASS, An information handling package from Newman College (1986), RM Nimbus version 2.0.

HALLIDAY, M. A. K. (1976), Theme and information in the English clause, in KRESS, G., *Halliday: system and function in language*, London: Oxford University Press, pp. 174-189.

HEDGES, W.D. (1984), The Data Base and Decision Making in Public Schools, *Computers in the Schools*, Vol. 1, N. 3, pp. 91-100.

HODSON, D. (1989), Using and constructing chemical databases - An example of conjectural CAL, *Unpublished paper.*

HOELSCHER, K. (1986), Computing and information: steering student learning, in OLDS, H. F., *The computer as an educational tool*, London: The Haworth Press, pp. 23-34.

HUBER, L. (1988), Computer Learning through Piaget's eyes, in JONES,A., *Computers in Education 5-13*, Open University Press, pp. 159-166.

HUNTER, B. (1983), Social Studies Education in the Information Society, *Social Education*, Vol. 47, N. 5, pp. 321-324.

HUNTER, B. (1987), Knowledge-Creative Learning with Data Bases, *Social Education*, Vol. 51, N. 1, pp. 38-43.

HUNTER, B., Social Studies Education in the Information Society, Social Education, Vol. 47, N. 5, May 1983, pp. 321-324.

INHELDER, B. & PIAGET, J. (1958), *The Growth of Logical Thinking fron Childhood to Adolescence*, London: Routledge and Kegan Paul, pp. 3-104.

JONASSEN, D. H. (1986), Improving Recall Using Database Management Systems: A Learning Strategy, *AEDS Journal,* Vol. 19, N. 2-3, pp. 109-123.

JONES, S. (1966), The effect of a negative qualifier in an instruction, *Journal of Verbal Learning and Verbal Behaviour*, Vol. 5, pp. 497-501.

KELLY, A.V. (1984), *Microcomputers and the Curriculum*, London: Harper & Row.

KENT, A. (1986), *The use of computers in the teaching of Geography*, London: ULIE.

KNIGHT, P. & TIMMINS, G. (1986), Using databases in history teaching, *Journal of Computer Assisted Learning*, 2, pp. 93-101.

LAKOFF, G. (1971), If's, and's and but's about conjuction, in FILLMORE, C. J. & LANGENDOEN, D. T., *Studies in linguistic semantics*, New York: Rinehart & Winston, pp. 115-149.

LANGHORNE, M. H. et al (1989), *Teaching with Computers: A New Menu for the '90s*, London: kogan Page, pp. 85-127.

LAWLER, R.W. (1985), *Computer Experience and Cognitive Development*, Chichester, Ellis Horwood.

MARCHIONINI, G. & SHNEIDEMAN, B. (1990), Finding Facts vs. Browsing Knowledge in Hypertext Systems, in BOY-BARRET, O. & SCANLON, E., *Computers and Learning*, Wokingham: Addison Wesley, pp. 250-261.

MARRAN, J. F., STOLTMAN, J. P. & LIOLIOS, N. (1986), School Geography and Computer Assisted Learning in the United States. From Simulations to Data Bases, in KENT, W.A. (Ed.), *The use of Computers in the Teaching of Geography*, London: ULIE, pp. 165-185.

MAYER, R. E. (1987), Cognitive aspects of learning and using a programming language, in CARROLL, J. M. (ed.), *Interfacing thought*, Cambridge, MA: MIT Press, pp. 61-79.

MAYNARD, K. (1988), Information Retrieval in Schools, *Information Technology and Learning*, Vol. 10, N. 3, pp. 63-64.

McCAWLEY, J. D. (1981), *Everything that linguistics have always wanted to know about logic but were ashamed to ask*, Chicago: University of Chicago Press, pp. 14-86.

McCLELLAND, J. (1986), A new twist on and old skill: Retrieving Information with computers to enhance decision-making processes, in OLDS, H. F., *The computer as an educational tool*, London: The Haworth Press, pp. 83-88.

McTIGHE, J. & LYMAN, F. (1988), Cueing thinking in the classroom: The promise of theory-embedded tools, *Educational Leadership*, 45, (7), pp. 18-24.

MENDRINOS, R. B. (1986), Uses of Data Bases in the Classroom, *Computers in the School*, Vol. 2, N. 4, pp. 65-70.

MITTLEFEHLDT, B. (1991), Social studies problem solving with data bases, *The computing teacher*, February, pp. 54-55.

MOONEN, J. & PLOMP, T. (1987), *Eurit 86: Developments in Educational Software and Courseware*, Oxford: Pergamon Press.

NATIONAL CURRICULUM COUNCIL (1989), *Technology 5 - 16 in the National Curriculum* (a report to the Secretary of State for Education and Science on the statutury consultation for attainment targets and programmes of study in technology), NCC, York, November.

NATIONAL CURRICULUM COUNCIL (1990), *Technology in the National Curriculum*, Department of Education and Science and the Welsh Office, pp. 45-49.

NATIONAL CURRICULUM COUNCIL (1989), *English in the National Curriculum*, Department of Education and Science and the Welsh Office, May.

OLDS, H. F. (1986), Information Management: A New Tool for a New Curriculum, in OLDS, H. F., *The computer as an educational tool*, London: The Haworth Press, pp. 7-22.

OSHERSON, D. N. & MARKMAN, E. (1975), Language and the ability to evaluate contradictions and tautologies, *Cognition*, Vol. 3, N. 3, pp. 213-226.

PAPERT, S. (1980), *Mindstorms: Children, Computers and Powerful Ideas*, Brighton: Harvester Press.

PARIS, S. G. (1973), Comprehension of Language Connectives and Propositional Logical Relationships, *Journal of Experimental Child Psychology*, Vol. 16, pp. 278-291.

PARKER, J. (1986), Tools for Thought, *The Computing Teacher* , October, pp. 21-23

PEA, R. D. & KURLAND, D. M. (1984), On the cognitive effects of learning computer programming, *New Ideas in Pshychology*, Vol. 2, pp. 137-168.

PERKINS, D.N. (1985), The Fingertip Effect: How Information-Processing Technology Shapes Thinking, *Educational Researcher,* Vol. 14, N. 7, pp. 11-17.

PHILLIPS, M. (1987), A progressive database, *IT and Learning*, Vol. 10, N. 1, pp. 12-13.

RAWITSCH, D. (1988), The Computerized Database: Not a Simple Solution, *The Computing Teacher*, Vol. 15, N. 14, pp. 34-37.

RAY, H. N. (1985), A study of the effect of different data models on casual users performance in writing database queries, *International Journal of Man-Machine Studies*, Vol. 23, pp. 249-262.

REISNER, P. (1977), Use of psychological experimentation as an aid to development of a query language, *IEEE Transactions on Software Engineering*, Vol. SE-3, N. 3, pp. 218-229.

REISNER, P., BOYCE, R. F. and CHAMBERLIN, D. D. (1975), Human factors evaluation of two data base query languages - Square and Sequel, in *Proceedings of the National Computer Conference* (Arlington, Virginia), New Jersey: AFIPS press, pp. 447-452.

RIGGS, H. N. (1990), Computer Data Base Activities in Upper Elementary Schools Social Studies: A Stimulus for Curriculum Integration, *History and Social Science Teacher*, Vol. 25, N. 3, pp. 145-150.

ROSS, A. (1984), Learning why to hypothesize: A case study of data processing in a primary school classroom, in KELLY, A. V., *Microcomputers and the Curriculum*, London: Harper and Row, pp. 64-83.

SCANLON, E. and O'SHEA, T. (1987), *Educational computing*, London: Open University.

SELF, J. (1985), *Microcomputers in Education. A Critical Evaluation of Educational Software*, Brighton, Harvester Press.

SENIOR, S. (1989), *Using IT across the National Curriculum*, Kent: Owlet Books, pp. 158-187.

SHNEIDERMAN, B. (1978), Improving the Human Factors Aspect of Database Interactions, *ACM Transactions on Database Systems*, Vol. 3, N. 4, pp. 417-439.

SHNEIDERMAN, B. (1980), *Software Psychology: Human Factors in Computer and Information Systems*, Cambridge, MA: Winthrop Publishers, pp. 143-214.

SMART, L. (1988), The database as a catalyst - a year with QUEST, *Journal of Computer Assisted Learning*, Vol. 4, N. 3, pp. 140-149.

SMYTH, M.M. & CLARK, S.E. (1986), My half-sister is a THOG: Strategic processes in a reasoning task, *British Journal of Psychology*, Vol. 77, pp. 275-287.

SPAVOLD, J, UNDERWOOD, J. & NEWMAN, I. (1990), Designing a Relational Datafile for Local History, *Computers and Education*, Vol. 15, N. 1-3, pp. 221-226.

SPAVOLD, J. (1989), Children and databases: An analysis of data entry and query formulation, *Journal of Computer Assisted Learning*, Vol. 5, pp. 145-160.

SPROWLS, R. C, (1975), Data base education for students of management, in *Proceedings of the National Computer Conference* (Virginia), New Jersey: AFIPS Press. pp. 907-911.

STAAL, J. F. (1968), AND, *Journal of Linguistics*, vol. 4, pp. 79-81.

STRAWSON, P. P. (1952), *Introduction to logical theory*, London, Methuen, pp. 2-101.

TABA, H. (1971), *Teacher's Handbook to Elementary Social Studies. An Inductive Approach*, London: Addison-Wesley.

TAPLIN, J.E., et al. (1974), Developmental Changes in Conditional Reasoning: Linguistic or Logical?, *Journal of Experimental Child Psychology*, Vol. 17, pp. 360-373.

TAYLOR (1980), *The Computer in the School: Tutor, Tool, Tutee,* New York: Teacher's college Press, Columbia University

THOMAS, J. C. (1977), Psychological issues in data base management, in *Proceedings of the InternationalConference on very large databases IV* (Tokyo), pp. 169-185.

THOMAS, J. C. and GOULD, J.D. (1975), A psychological study of query by example, in *Proceedings of the National Computer Confere*nce (Arlington, Virginia), New Jersey: AFIPS press, pp. 439-445.

THOMAS, R. (1988), The Student-Designed Database, *The Computing Teacher*, Vol. 15, N. 5, pp. 17-19.

TRABERMAN, T. (1984), Using Microcomputers to Teach Global Studies, *Social Education*, Vol. 48, February, pp. 130-137.

UNDERWOOD, J. & UNDERWOOD, G. (1987), Data Organisation and Retrieval by Children, *British Journal of Educational Psychology*, Vol. 57, pp. 313-329.

UNDERWOOD, J. (1986), The role of the computer in developing childrens' classificatory abilities, *Computers and Education*, Vol. 10, N. 1, pp. 175-180.

UNDERWOOD, J. (1988), An investigation of teacher intents and classroom outcomes in the use of information-handling packages, *Computers in Education*, Vol. 12, pp. 91-100.

UNDERWOOD, J. D. M. & UNDERWOOD,G. (1990), *Computers and Learning. Helping children acquire thinking skills*, Oxford: Basil blackwell.

UNDERWOOD, J. D. M. (1985), Cognitive demand and CAL, in REID, I. & RUSHTON,J., *Teachers, computers and the classroom*, Manchester, Manchester University Press, 1985, pp. 25-37.

UNDERWOOD, J., SPAVOLD, J. & UNDERWOOD, G. (1990), Novice use of a relational database: A case study of a local history data file, *Computers and Education*, Vol. 15, N. 1-3, pp. 227-232.

UNDERWOOD, J., SPAVOLD, J. & UNDERWOOD,G. (1990), Learning to use a relational database, in ESTES, N et al. (Eds.), *The seventh international conference on technology and education: new pathways to learning through educational technology*, Edinburgh: CEP consultants, pp. 286-289.

VAN DIJK, T. A. (1989), *Text and context; Explorations in the semantics and pragmatics of discourse*, London: Longman, pp. 43-129.

VOCKELL, E.L., & KOPENEC, D. (1989), Choosing the Right Database Management Program, *The Computing Teacher,* August/September, pp.29-43.

WARNER, M.(1988), Developing Databases Files for Student Use, *The Computing Teacher,* Vol.15, N.7, pp.44-47.

WASON, P. C. (1962), *Psychological Aspects of Negation: An experimental enquiry and some practical applications*, London: Communication Research Centre, University College London.

WATSON, D.(1984), Microcomputers in secondary education - a perspective with particular reference to the humanities, in KELLY, A. V., *op. cit.*, pp. 125-144.

WATSON, J. & STRUDLER, N. (1989), Teaching Higher Order Thinking Skills with Databases, *The Computing Teacher*, Vol. 16, N. 4, pp. 47-50.

WATSON, J. (1988), Database Activities in a One-Computer Classroom, *The Computing Teacher*, August/September, pp. 21-23.

WEIZENBAUM, J. (1976), *Computer Power and Human Reason: From Judgement to Calculation*, San Francisco: Freeman.

WELFORD, R. (1989), Introductory Analysis in the Social Sciences Using Small-Scale Databases, *Journal of Computer Assisted Learning*, Vol. 5, pp. 95-102.

WHEELER, F. (1987), The New Ready-Made Databases: What They Offer Your Classroom, *Classroom Computer Learning*, Vol. 7, N. 6, pp. 28-32.

WHITE, C.S. (1987), Developing Information-Processing Skills through Structured Activities with a Computerized File-management Program, *Journal of Educational Computing Research*, Vol. 3, N. 3, pp. 355-375.

WIEBE, J. H. (1990), Teaching Mathematics with Technology: Data Base Programs in the Mathematics Classroom, *Arithmetic Teacher*, Vol. 37, N. 5, pp. 38-40.

WOOD, D. J. (1988), *How Children Think and Learn,* Oxford: Basil Blackwell.

WORSTER, B & MORRISON, D.M. (1986), Using Appleworks in the Geography Class: From Transparent Maps to Sortable Electric Data, *Computers in the School,* Vol. 3, N. 1, pp. 63-73.

APPENDIX 1

Bezanilla, M. J. and Ogborn J. (1992), Logical sentences and searches, *Journal of Computer Assisted Learning*, 8, 37-48.

Please note pages 211-222 contain copyrighted material (a published journal article) and have been redacted.

The full text article can be found here - http://dx.doi.org/10.1111/j.1365-2729.1992.tb00383.x

APPENDIX 2: GRASS and POP

Section A: POP data file - Field description

Section B: GRASS (Graphics Searching and Sorting) - Main Interrogation Menu

Appendix 2

Section A

POP data file - field explanation

COUNTRY	Name of country
CONTINENT	Name of continent
POPULATION	in millions 1984
GNP/PERSON	gross national product per person in US dollars in 1984
LIFELENGTH	average life expectancy at birth in years in 1984
POPGROWTH	average annual growth of population to double in size if the present growth rate continues
DOUBLETIME	time in years for the population to double in size if the present growth rate continues
C.BIRTH R	crude birth rate per 1000 of the population in 1984
C.DEATH R	crude death rate per 1000 of the population in 1984
CHANGE CBR	percentage change in crude birth rate 1975-84
CHANGE CDR	percentage change in crude death rate 1975-84
CONTRACEPT	percentage of married women of childbearing age who are using contraception
INFANTMORT	the infant mortality rate, which is the number of infants who die before reaching one year of age, per 1000 live births in 1984
FOODPRODTN	the index of food production per person which shows the average annual quantity of food produced in 1982-84 compared with 1974-76. The index each country for 1974-76 is 100
POP/DOCTOR	average number of people per doctor in 1981
FPLANNING	level of government support
SECONDEDUC	numbers attending secondary schools as a percentage of the 12-17 age group
URBANPOP	the percentage of the population living in urban areas

Appendix 2

Section B

GRASS (Graphics searching and sorting) - *Newman College (1985)*

INTERROGATION MENU

1. Display records
2. Search data
3. Sort data
4. Graphs and figures
5. Save records
6. Teacher's page
7. Change data file
8. Finish the program

1. Display records

This option enables details of records to be shown on the screen. Each field is selected by moving a shaded box over the menu required and pressing SPACE. After having selected all the fields that are wished to be displayed, "Display" must be selected from the menu.

2. Search data

a. You select the field

b. You select the "relationship" or "test" for the search. These are different for numeric fields and alphanumeric ones.

For numeric fields:

. is equal to
. is not equal to
. is greater than
. is less than

For alphanumeric fields:

. includes
. does not include
. is the same as
. is not the same as
. comes before (Numerical characters come before letters.
. comes after Letters are in alphabetical order)

c. You type in the value. Then, RETURN.

d. You are asked to confirm your search.

At any stage, you can press the ESCAPE key to interrupt the search and return to the main menu.

e. Then, you may:

- . see the results of your search with the display records option
- . look at graphs or figures for that search
- . add to the search with an OR condition. The question will be met if either one OR the other condition is met
- . narrow the records with an AND condition. The question will only be met if both conditions are met
- . save the records that fulfil the question
- . sort the records for this question
- . end the search with the "Start a new search" condition

3. Sort data

Records can be sorted at any time. Sorting can be on either numeric or alphanumeric fields.

- . on numeric fields with the "biggest" or "smallest" first
- . on alphabetic fields into alphabetical order

4. Graphs and figures

The range of graphs or figures that are available depends on the type of field that is selected.

- . For numeric data:
  - . Histogram
  - . Cumulative graph
  - . Scattergraph
  - . Average or mean
  - . Median
- . For alphanumeric data:
  - . Pie chart
  - . Count graph

5. Save records

Some or all of the records may be saved at any time. Fields selection is by ticking the filed names in the same manner as in "display records". The file name for the new file must be less than eight characters. Once saved, the new file can be loaded at any time.

6. Teacher's page

The teacher may set up certain features of GRASS via the teacher's page. The options are selected by moving the highlight box with the cursor keys, and using the SPACE BAR to toggle between "Yes" and "No" until all the features are set up as required. By default all features are set to "No".

7. Change data file

This option allows the user to change one data file for another at any time. The program returns to the display of data files that may be chosen from the current disk.

8. Finish the program

This is the correct method to exit the GRASS query program. To completely leave GRASS, "Finish using the GRASS package" from the main menu must be chosen.

APPENDIX 3: Instruments of Data Collection (*)

Section A: Exercises 1 and Exercises 2
Section B: "Report editing" task
Section C: Logic (And/Or) questionnaires (Interpreting and Constructing tasks)
Section D: Questionnaire on children's understanding of the nature of data
Section E: Questionnaire on children's understanding of the nature of data search and structure

(*) Raw data of the instruments are available from the author

Appendix 3

Section A - Exercises 1 and Exercises 2 -

NAME:	SCHOOL:
AGE:	LEVEL:
	DATE:

EXERCISES 1

Answer the following questions and explain what you did to get the answers:

1. What is the population of Japan?

2. Which is the country with the lowest death rate?

3. In how many countries is the population smaller than 3 million?

4. Do countries with a high GNP have a large percentage of people living in towns and cities?

5. Which are the African countries whose growth of population is higher than 2.5%?

NAME: SCHOOL:
AGE: LEVEL:
DATE:

EXERCISES 2

Answer the following questions and explain what you did to get the answers:

1. Which are the three most populated countries?

2. In which continent is Benin?

3. Is there any relationship between birth and death rates?

4. How many countries have a birth rate under 15 per 1000 and a death rate under 11 per 1000?

5. How many countries belong to Europe?

Appendix 3 - Section B: "Report editing" task

NAME: SCHOOL:

AGE: LEVEL:

DATE:

In the Geography Department, we have found a Report titled "Population of the World", which contains particular information about different countries. It seems that some of the information contained in it is true, and some is false. We want you to check this report by using the data base and find out which statements are true and which are false.

Number each statement you check, and give it a √ (true) or X (false).
On the next page, explain how you checked each statement using the computer.

Report: "Population of the World"

The population of the world is very unevenly distributed. Some of the countries are sparsely populated, such as Nicaragua, New Zealand, Singapore and Jordan, with less than 4 million inhabitants. Some others are over populated, like India, which has more than 1000 million people. Besides, the world population is not expanding at the same rate everywhere. In India, for instance, the population growth rate is high, 2.3 % per year, and this is made worse by the fact that its population is already the largest of the world.

In some countries, the need to reduce the population growth rates has become urgent, and the reduction of birth rates seems to be the only means of doing it. The birth rate is over 50 in more than half of the African countries, while in European countries, for instance, the birth rates are very low which may be because there is a high percentage of women using contraceptives.

The Gross National Product, that is, the average income of a country, gives an idea of its development. How well off countries are is reflected by the GNP. For example, USA, which is considered as the most developed country in the world, has the highest GNP per person, and on the contrary, the GNP in the African countries is not higher than 750S, if we exclude Mauritius, Tunisia and Zimbabwe.

Less developed areas, such as the three African countries Sudan, Uganda and Bangladesh, have many interesting features. For example, in countries with low GNP, the infant mortality rates are lower than in rich countries.

Statements checked

1

2

3

4

5

6

7

8

9

10

11

12

Appendix 3

Section C: Logic questionnaires (AND / OR / NOT)

Interpreting task

In the Electronics Department of a very big store, they have shelves which contain leaflets with information about different products such as TVs, cameras, and CD players.

The leaflets are spread out on the shelves regarding the type of product and its country of origin, in the following way:

	European	Japanese	American
TV			
CD player			
camera			

Your job is to tick the shelves you should look on in order to find some particular leaflets, and to put a cross on the rest of the shelves.

For example, if we want leaflets about TVs, regardless where they were made, we should look on the following ticked shelves, and not on the crossed ones:

	European	Japanese	American
TV	✓	✓	✓
CD player	X	X	X
camera	X	X	X

If we want leaflets about cameras which are not European, we should look on the following shelves ticked below:

	European	Japanese	American
TV	X	X	X
CD player	X	X	X
camera	X	✓	✓

Tick the shelves to look on
if you want leaflets about:

goods which are CD players
AND
which are Japanese

Put a cross on the remaining shelves

	European	Japanese	American
TV			
CD player			
camera			

Tick the shelves to look on
if you want leaflets about:

goods which are not TVs
AND
which are American

	European	Japanese	American
TV			
CD player			
camera			

Tick the shelves to look on
if you want leaflets about:

goods which are not cameras
OR
which are Japanese, or both

	European	Japanese	American
TV			
CD player			
camera			

Tick the shelves to look on
if you want leaflets about:

goods which are TVs
OR
which are not TVs

	European	Japanese	American
TV			
CD player			
camera			

Tick the shelves to look on
if you want leaflets about:

goods which are not CD players
OR
which are not Japanese, or those which satisfy both criteria

	European	Japanese	American
TV			
CD player			
camera			

Tick the shelves to look on
if you want leaflets about:

goods which are cameras
AND
which are not cameras

	European	Japanese	American
TV			
CD player			
camera			

Tick the shelves to look on
if you want leaflets about:

goods which are not CD players
AND
which are not European

	European	Japanese	American
TV			
CD player			
camera			

Tick the shelves to look on
if you want leaflets about:

goods which are cameras
OR
which are Japanese, or both

	European	Japanese	American
TV			
CD player			
camera			

Appendix 3 - Section C - Constructing task

In the Police Office, they have shelves which contain papers with the passport details of different people: men, women and children.

These papers are organised on the shelves regarding whether those people are men, women or children, and their nationality: English, Scottish or Welsh, in the following way:

	English	Scottish	Welsh
Women			
Men			
Children			

We are going to show you eight different ways in which those papers can be spread out on the shelves, and the following sets of alternatives:

ARE	English	OR	ARE	English
	Scottish			Scottish
ARE NOT	Welsh	AND	ARE NOT	Welsh
	women			women
	men			men
	children			children

Your job is to circle those alternatives (<u>one alternative in each column</u>), which allow you to describe the people whose passport details are shown on the shelves:
For example, if the papers are spread out on the following shelves:

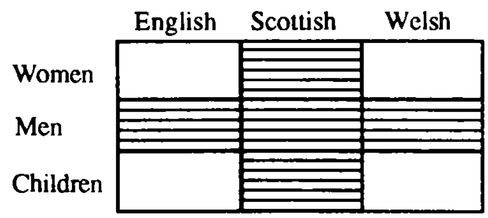
	English	Scottish	Welsh
Women			
Men			
Children			

We can say, by circling the right alternatives, that the people whose passport details are shown on the shelves are women and also are not Welsh:

ARE	English	OR	ARE	English
	Scottish			~~Scottish~~
ARE NOT	Welsh	AND	ARE NOT	Welsh
	women			~~women~~
	~~men~~			men
	children			children

Circle one alternative in each column,
so as to make a statement which is true:

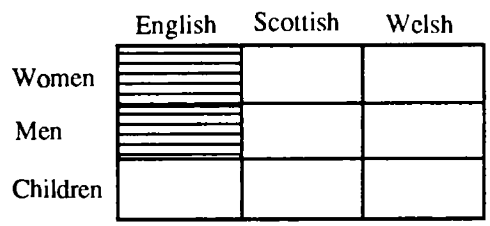
All the people, whose passport details
are shown on the shelves:



ARE	English	OR	ARE	English
	Scottish			Scottish
ARE NOT	Welsh	AND	ARE NOT	Welsh
	women			women
	men			men
	children			children

Circle one alternative in each column,
so as to make a statement which is true:

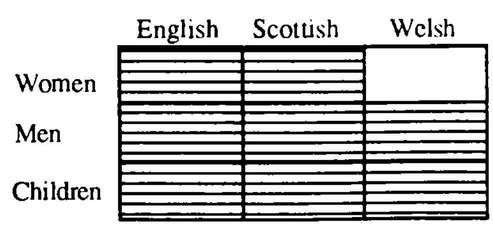
All the people, whose passport details
are shown on the shelves:



ARE	English	OR	ARE	English
	Scottish			Scottish
ARE NOT	Welsh	AND	ARE NOT	Welsh
	women			women
	men			men
	children			children

Circle one alternative in each column,
so as to make a statement which is true:

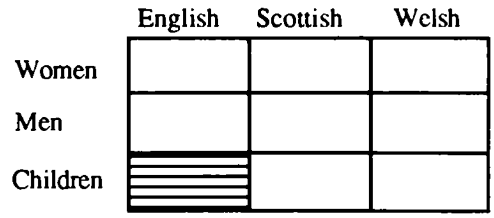
All the people, whose passport details
are shown on the shelves:



ARE	English	OR	ARE	English
	Scottish			Scottish
ARE NOT	Welsh	AND	ARE NOT	Welsh
	women			women
	men			men
	children			children

Circle one alternative in each column, so as to make a statement which is true:

All the people, whose passport details are shown on the shelves:



ARE
ARE NOT

English
Scottish
Welsh
women
men
children

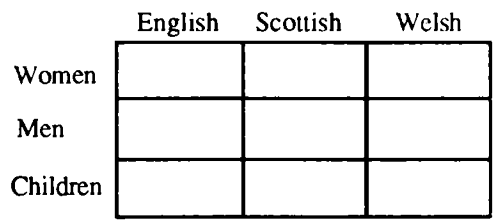
OR
AND

ARE
ARE NOT

English
Scottish
Welsh
women
men
children

Circle one alternative in each column, so as to make a statement which is true:

All the people, whose passport details are shown on the shelves:



ARE
ARE NOT

English
Scottish
Welsh
women
men
children

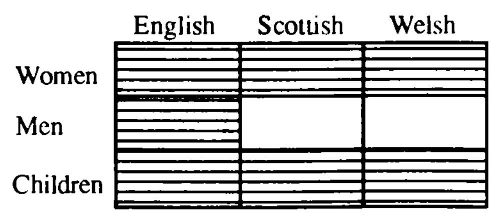
OR
AND

ARE
ARE NOT

English
Scottish
Welsh
women
men
children

Circle one alternative in each column, so as to make a statement which is true:

All the people, whose passport details are shown on the shelves:



ARE
ARE NOT

English
Scottish
Welsh
women
men
children

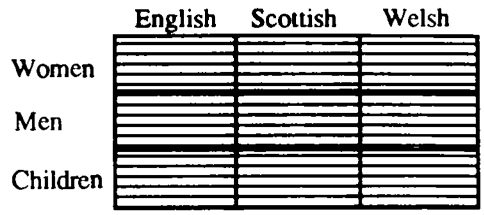
OR
AND

ARE
ARE NOT

English
Scottish
Welsh
women
men
children

Circle one alternative in each column,
so as to make a statement which is true:

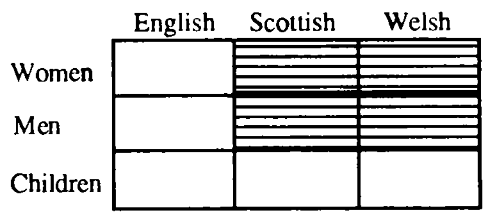
All the people, whose passport details
are shown on the shelves:



ARE	English	OR	ARE	English
	Scottish			Scottish
ARE NOT	Welsh	AND	ARE NOT	Welsh
	women			women
	men			men
	children			children

Circle one alternative in each column,
so as to make a statement which is true:

All the people, whose passport details
are shown on the shelves:



ARE	English	OR	ARE	English
	Scottish			Scottish
ARE NOT	Welsh	AND	ARE NOT	Welsh
	women			women
	men			men
	children			children

Appendix 3

Section D

Questionnaire on children's understanding of the nature of data

NAME SCHOOL

AGE LEVEL

DATE

For each of the questions below, say if you think it could or it could not be answered by retrieving data from a data base.

Tick the box in the following way:

[√] Could

[X] Could not

1. Who is the Prime Minister of the British Government? [ ]
2. Will it rain more next year than it did this year? [ ]
3. Which is the most objective British newspaper? [ ]
4. How many cars were bought in England in 1988? [ ]
5. Is the Soviet Union serious about arms reductions? [ ]
6. Which is the title of the first book written by Agatha Christie? [ ]
7. Do children enjoy using the computer? [ ]
8. Which is the highest mountain in North America? [ ]
9. What would happen to Sweden's population if there was a sudden baby boom in the next five years? [ ]
10. Is Margaret Thatcher going to be reelected in the next elections? [ ]
11. Is it more convenient to go to Bristol by coach or by train? [ ]
12. Which is the biggest country of the EEC? [ ]
13. Do football fans not really mean any harm? [ ]
14. What is the relationship between the production rate of grapes and the price of the wine? [ ]
15. Do girls like getting flowers as a present? [ ]
16. Which was the best directed film last year? [ ]
17. Do richer countries have more people studying at Universities? [ ]
18. Do people wear smart clothes at a party? [ ]
19. Which was the warmest month in Britain last year? [ ]
20. Is there a link between the number of inhabitants in a town and the number of schools? [ ]

Appendix 3 - Section E

Questionnaire on children's understanding of the nature of data search and structure

Suppose you have cards with information about children's books. Each card says:

- Title of the book
- Kind of book (fairy-tale, textbook, fable, short-story...)
- Name of the author/s
- Date of publishing
- Publishing firm
- Number of pages
- Number of chapters/units
- Number of illustrations
- Type of binding
- Size
- Price

The information on these cards is put into a database on a computer.

We are going to tell the computer to find some information about the books and try to imagine how it might do it.

• Suppose we tell the computer to find the name of the author of a book titled "Jane Eyre". How do you imagine it might try to do it?

• Suppose we tell the computer to find the titles of all the fairy-tales. How do you imagine it might try to do it?

• Suppose we tell the computer to find all the textbooks which are cheaper than £5. What do you imagine it will do to try to find them?

• Suppose we ask the computer to put the books in order according to their price, ranging from the cheapest to the most expensive. How do you imagine it might try to do such a thing?

• Suppose we tell the computer to find out if the most expensive books usually have the most pages. How do you imagine it might try to do it?

Write down here any other questions about the books which you think the computer could answer.

APPENDIX 4: Training on GRASS

Appendix 4

TRAINING ON GRASS

The training consisted of three main parts:

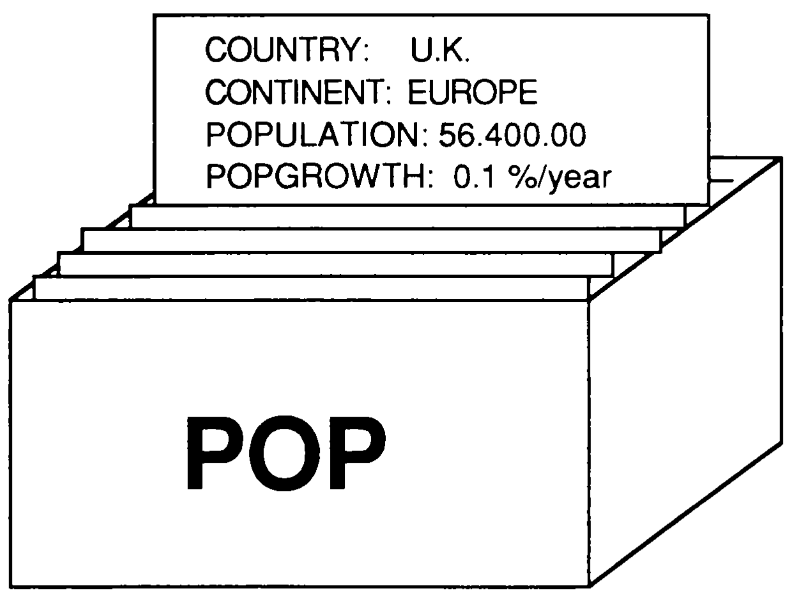
1. Children were taught what a database is very briefly, and introduced to the notions of "record" and "field".

2. Children were introduced to some essential information about the system, like how to start the program, and the function of the arrow keys, Space Bar, Return and Escape, as well as the screen layout.

3. Children were taught how to interrogate the POP data file with GRASS. Each option of the Menu was explained carefully, and examples in form of questions were presented and solved step by step in interaction with the system.

1. WHAT A DATA BASE IS

A data base is a collection of information about a particular topic that is held on a disc. This information is much the same as that on a card in a filing cabinet. For instance, a data file called POP has information about many countries of the world. On a card, the information would look like this:



On a data base, the information has to be carefully structured.

In the POP data file, information about each of the 65 countries is called a **record**. Each record is the equivalent of one card in the filing cabinet.
In each record, there are 18 different items of information. Each item is called a **field**.

This means that for the whole data file there are 65 x 18 = 1170 different individual items of information. Because this information is stored on the disc, any items or combination of items can be recalled immediately.

2. ESSENTIAL INFORMATION ABOUT GRASS

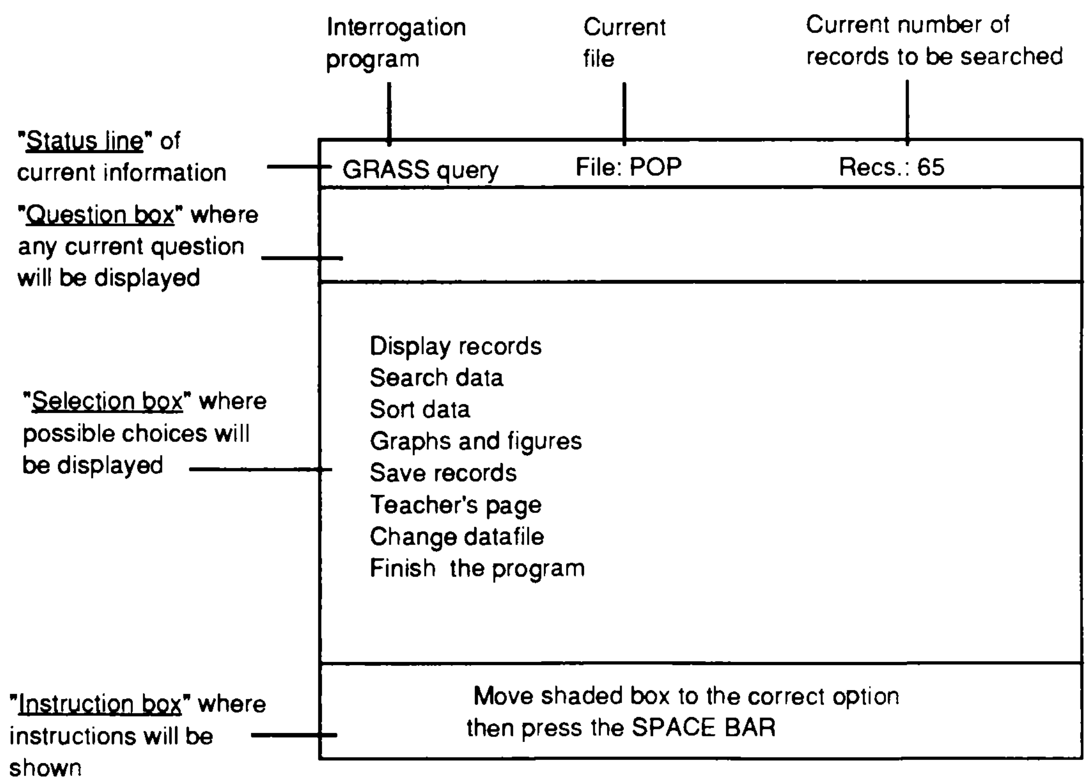
Children were given the following indications, which were written down on the whiteboard:

1. Use SHIFT BREAK to start the program

2. The method of user control throughout the program is as follows:

    a. The arrow keys move the red box around the screen
    b. The SPACE BAR is used to select an option that is highlighted
    c. Any input typed from the keyboard must be ended with RETURN
    d. ESCAPE returns you to the main menu

3. To explain the Screen Layout, I made a representation of it on the whiteboard as follows:



Each box was explained briefly in a very simple way.

3. INTERROGATION WITH GRASS

The four first options of the Main Menu (Display records, Search data, Sort data and Graphs and Figures) were explained carefully, with the aid of examples. I was presenting questions and solving them step by step on the whiteboard, and children were doing the same interacting with the system.

1. Display records
This option enables details of records to be shown on the screen

Select each field by moving the shaded box over the name you require and pressing SPACE. An asterisk (*) will appear to the left of the name to mark it. In this way several fields can be identified for the display.

The SPACE BAR acts as a "toggle" switch. Pressing it once displays an asterisk by that name. A second press will remove it. To select all the fields use "All fields" from the menu.

When you have selected all the fields you wish to display, move the highlighted box and use the SPACE BAR to select "Display".

Fields will be displayed across the screen if there is room, otherwise they will be listed down the screen. Records will be shown one at a time. For long records, the display will be given in two or more parts. Pressing the SPACE BAR will give the next record, or part of record. Pressing ESCAPE will interrupt the display and return to the menu.

Now, you can display all the information for all the countries:

> *Display records*
> ** All fields*
> *Display*

Or you can display part of the information for all the countries:
For example,
what would you do if you wanted to know the population in each country?

> *Display records*
> ** COUNTRY*
> ** POPULATION*
> *Display*

And, if you wanted to know the GNP per person, the birth rate, the death rate, and the length of life of each country?

> *Display records*
> ** COUNTRY*
> ** GNP/PERSON*
> ** C.BIRTH R.*
> ** C.DEATH R.*
> ** LIFELENGTH*
> *Display*

2. Search data
To search data you do not have to remember any commands or the field names because you will be prompted on the screen.

Select the field with the highlight box and SPACE BAR and the field name is entered in the "question box" for you. The whole question for your search will be built up in the "question box" stage by stage.

The "relationship" or "test" for the search must now be selected. These are different for numeric fields (such as HEIGHT) and alphanumeric ones (such as NAME).

For numeric fields:
- is equal to
- is not equal to
- is greater than
- is less than

For alphanumeric fields:
- includes
- does not include
- is the same as
- is not the same as
- comes before
- comes after

Finally, the "value" is typed in from the keyboard. This value must be followed by RETURN.

After a search has been entered you are asked to confirm it before the file is searched. A negative answer to the question "is this all correct?" will give you the opportunity to enter the whole question again. But at any stage you can press ESCAPE to interrupt the search and return to the main menu.

Once a search has been made the question remains in the question box, because it is still active. If you wish to remove it, you must end this search by the option in the menu to "Start a new search".

With a question active you may:
a) See the results of your search with the display records option
b) Look at graphs or figures for that search
c) Add to the search with an OR condition
d) Narrow down the search with an AND condition
e) Save the records that fulfil this question
f) Sort the records for this question
g) End the search with the "Start s new search" option

Now, in order to know what is the population of Canada, we can do the following:

Search data
** COUNTRY*
same as
(You type in) Canada
Display records
** COUNTRY*
** POPULATION*
Display

Or, to know in which countries the population is higher than 35 million, we can proceed as follows:

Search data
** POPULATION*

is greater than
(You type in) 35
Display records
** COUNTRY*
** POPULATION*
Display

Complex Searches with GRASS

The **OR** option allows you to add to the search, because the question will be met if either one OR the other condition is met, thus increasing the number of records selected.

The **AND** option allows you to narrow down the search, because the question will only be met if both conditions are met, thus reducing the number of records.

Once you have confirmed a search condition, you cannot change it. It may only be removed when you "Start a new search".

For instance, suppose you want to find out the name of the countries which are European AND whose length of life is higher than 75, we can do the following:

Search data
** CONTINENT*
same as
(You type in) Europe
Narrow down the search (AND)
** LIFELENGTH*
is greater than
(You type in) 75
Display records
** COUNTRY*
** CONTINENT*
** LIFELENGTH*
Display

We find out that out of the 14 European countries, 6 have a length of life higher than 75.

On the contrary, suppose you ask the computer for all those countries which are European or which have a length of life higher than 75. The query could be developed in the following way:

Search data
** CONTINENT*
same as
(You type in) Europe
Add to the search (OR)
** LIFELENGTH*
is greater than
(You type in) 75
Display records
** COUNTRY*
** CONTINENT*
** LIFELENGTH*
Display

And you can realise that the number of records has increased to 19. This means that there are 14 countries which belong to Europe, and 5 that though not being European have a length of life higher than 75 years.

3. Sort data

Records can be sorted at any time with GRASS. Sorting can be on either numeric or alphabetical fields. Numeric values can be sorted with either "biggest first" or "smallest first", while alphabetic records will always be sorted into alphabetical order.

For instance, suppose you want to know which are the five most populated countries. You can use the sort data option in the following way:

Sort data
** POPULATION*
biggest first
Display records
** COUNTRY*
** POPULATION*
Display

In the same way, suppose you want to know what is the lowest figure for infant mortality, you can sort data on this field in the following way:

Sort data
** INFANTMORT*
smallest first
Display records
** INFANTMORT*
Display

Once sorted the data is held in that order until sorted on another field. This can be useful because it allows data to be sorted on more than one field.

4. Graphs and figures

The range of graphs or figures that are available depends on the type of field is selected. A histogram, cumulative graph or scatter graph may only be drawn using numeric data. The average and median are also available for numeric data. For alphanumeric data, pie charts and count graphs can be displayed.

The graphs allow you to see the data in different ways.

You can try, for instance to display the information about population change in all the countries on a histogram.

Graphs and figures
POPGROWTH
Histogram

The histogram shows the value divided into classes for any field that contains numbers.

Press a letter to see what each bar represents.

The scattergraph in "Graphs and figures" gives a useful guide to the relationships between values in different fields.

For instance, try to see what kind of relationship there is between the GNP of a country and the average number of people per doctor.

Graphs and figures
** GNP/PERSON*
Scattergraph
** URBANPOP*

The pie chart can be used to display data that is not in number form.
For instance:

Graphs and figures
** F.PLANNING*
pie chart

APPENDIX 5: Factor Analysis

13 year olds - Exercises

Factor Extraction Method: Principal Component Analysis
Transformation Method: Orthotran / Varimax
Number of Factors: 5
Factor Scores: Oblique Solution
Bartlett Test of Sphericity DF: 54 Chi Square: 120.292 p: .0001
Factors correlations (r < .45)

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
v 1.1	-.304	-.159	**.928**	.069	-.002
v 1.2	.003	**.952**	-.002	-.027	-.021
v 1.3	**.867**	-.221	-.343	-.159	.106
v 1.4	-.194	1.039	-.126	.004	.032
v 2.1	.033	.007	-.008	.022	**.95**
v 1.1b	**.548**	.099	**.707**	-.191	.02
v 1.2b	.129	.049	-.026	**.777**	-.327
v 1.3b	-.22	-.078	-.016	**.992**	.23
v 1.4b	**.631**	.018	.14	.085	.153
v 2.1b	**.831**	-3.62E-4	-.052	-.001	-.173

13 year olds - Report

Factor Extraction Method: Principal Component Analysis
Transformation Method: Orthotran / Varimax
Number of Factors: 3
Factor Scores: Oblique Solution
Bartlett Test of Sphericity DF: 44 Chi Square: 69.129 p: .0091
Factors correlations (r < .47)

	Factor 1	Factor 2	Factor 3
v 1.1	**.54**	-.132	.262
v 1.2	**.921**	.264	-.082
v 1.3	**-.634**	**-.878**	**.571**
v 1.4	**.88**	-.018	-.032
v 2.1	3.170E-4	3.701E-4	**.84**
v 1.1b	.182	**.952**	.306
v 1.3b	.195	**-.52**	.42
v 1.4b	**.698**	.303	.004
v 2.1b	.402	.057	**.51**

16 year olds - Exercises

Factor Extraction Method: Principal Component Analysis
Transformation Method: Orthotran / Varimax
Number of Factors: 3
Factor Scores: Orthogonal Solution
Bartlett Test of Sphericity DF: 54 Chi Square: 98.736 p: .0002
Factors correlations (r < .20)

	Factor 1	Factor 2	Factor 3
v 1.1	**.932**	.145	.133
v 1.2	-.128	**.652**	-.197
v 1.3	.378	.176	-.054
v 1.4	-.183	-.175	**.553**
v 2.1	.229	.013	**.863**
v 1.1b	**.636**	-.15	-.011
v 1.2b	.084	**.89**	.066
v 1.3b	.159	**.825**	-.043
v 1.4b	**.747**	-.145	-.344
v 2.1b	**.783**	.152	.353

16 year olds - Report

Factor Extraction Method: Principal Component Analysis
Transformation Method: Orthotran / Varimax
Number of Factors: 4
Factor Scores: Orthogonal Solution
Bartlett Test of Sphericity DF: 44 Chi Square: 64.479 p: .0237
Factors correlations (r < .20)

	Factor 1	Factor 2	Factor 3	Factor 4
v 1.1	**.802**	.329	.181	-.124
v 1.2	.042	.019	**.899**	-.11
v 1.3	**.73**	-.088	.312	.295
v 1.4	-.084	**.874**	-.219	.094
v 2.1	**.659**	-.114	-.09	.052
v 1.1b	**.668**	.109	-.47	-.326
.1.3b	.087	.126	-.078	**.922**
v 1.4b	**.621**	.44	-.142	.328
v 2.1b	.16	**.781**	.28	.046