PROBLEMS IN DEFINING AND ELICITING 'SCIENTIFIC' PROCESSES USING PRACTICAL TASKS WITH PRIMARY SCHOOL CHILDREN.

ANGEL DANIEL LOPEZ Y MOTA

Thesis submitted for degree of
PhD in Science Education at
the Institute of Education,
University of London, 1991.

ABSTRACT

The inclusion of science in the primary school curriculum makes it necessary to have plausible scientific activities that fulfil certain educational objectives. The purpose of this study is to show to what extent it is possible to combine such criteria.

Tasks have been constructed, each in two versions - Structured and Investigation plus Goal - in each four topic domains. Levels of success are given for both versions of the tasks, both overall and in relation to their cognitive demands.

The tasks are constructed so as to elicit, as far as possible, a number of 'scientific processes'. This notion is not, however, taken for granted. The study looks at processes in the context of the tasks; at when it makes sense to label processes as such or when they are better considered as content bound, and at the nature of what in fact has been elicited from children.

It has been possible to answer some questions related to children's 'scientific' behaviour such as, How complete are children's investigations?, Do children notice relevant phenomena? Do they draw conclusions from what they noticed? How well do children identify and control variables?, How good are children at using 'what-if reasoning', What explanations of the phenomena do children give? and, Do children make notes when doing an investigation?

The framework of the study describes and compares two models of defining and eliciting 'scientific' processes, leading to the organization of the literature review in terms of: problems of transfer from the nature of science, problems of defining and matching, and problems of eliciting and discriminating. The conclusions are organized around three main areas of concern: 1) the tasks, 2) the processes, and 3) children's behaviour. The first comments on the tasks as potentially pedagogic and diagnostic devices. The second considers the problems of defining and eliciting 'scientific' processes. The last recapitulates the findings on children's behaviour, emphasizing some of their commonsense features.

To Lupita, Ricardo and Andrea

ACKNOWLEDGEMENTS.

I thank all those, supervisors, colleagues, technicians and friends, who have helped in one way or another to make this thesis possible; especially Jon Ogborn who drove me into living with uncertainty, and the National Council for Science and Technology in Mexico for having sponsored my studies.

TABLE OF CONTENTS:

INDEX OF FIGURES.

INDEX OF TABLES.

Chapter 1.
A FRAMEWORK FOR THE STUDY.

This chapter discusses some of the problems concerning so called 'scientific processes', especially those concerning primary school children, in order to establish a framework for the research.

Interest in 'scientific processes', as a framework within which to plan and think about the meaning of 'science' for primary school science activities, has been and continues to be strong. There has been a continuing debate between the importance of 'process' and 'content', the latter usually understood as acquiring scientific concepts. The particular way in which 'scientific processes' have been understood by different curriculum development projects and researchers has certainly been varied (see chapter 2). A common approach has been to identify 'scientific processes' from the philosophy of science, whether in the context of justification or of discovery, and then to derive from this their importance for primary school children. The main feature of this approach is that the possibility of the existence of the chosen 'science processes' in children, given suitable teaching, is in a sense taken for granted, due to the way they are defined: indeed they define what is taken to be important for children to learn. The logic of this position is essentially linear, as illustrated in Figure 1.1: pre-defined 'processes' are to be taught to children and then to be elicited from them in some valid and reliable way. This way of proceeding tends to assume that there is nothing essentially problematic about defining 'scientific processes' in one context (science and scientists) and looking for them, or something like them, in another (the child and the school). What the child can or should do need not be the mirror image of what a scientist does.

In this research I shall use a different viewpoint, shown in Figure 1.2, in which the nature of and the connections between the same elements - the nature of science, processes, tasks to elicit processes, and measures of performance - are all able to be taken as problematic. That is, I shall adopt a view in which the research is addressed, via empirical results, to a critical analysis of the concept of 'scientific processes in children'. But the study is still focused on processes, so that the learning of scientific concepts is not directly addressed, without implying that problems coming from the interaction of conceptual and process aspects do not exist - indeed a number of these will be considered.

1.1 A LINEAR MODEL.

This extreme and perhaps unrealistic model (Figure 1.1), represents as a logical sequence the derivation from a source which defines the nature of 'scientific processes', through an instrument to elicit such behaviours, and measurement of performance of children. Seen as such it discourages us from questioning the relationships between all these elements.

In this model the purpose of engaging children in 'scientific' activities, whatever this means, tends to be seen as evaluating or assessing how well they do in such activities, in order to judge the attainment of some educational objectives. Two elements that stand out in this model are the test instrument and its function of summary assessment.

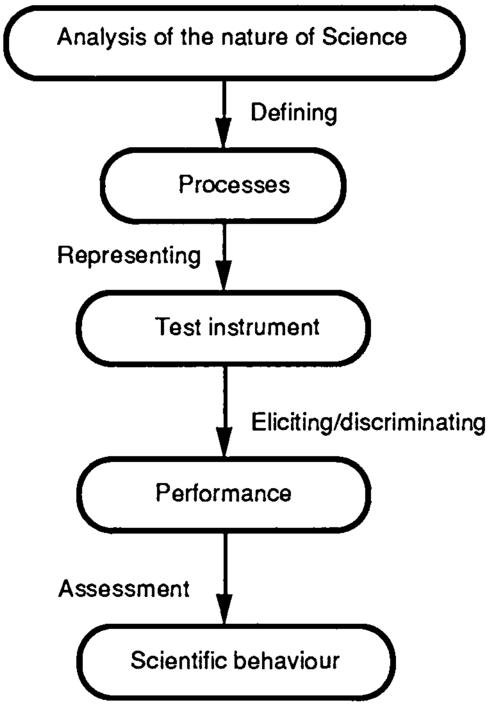


Figure 1.1: linear model of defining and eliciting children's scientific behaviour.

The model starts with a source of reference or inspiration, which allow 'scientific processes' to be drawn from an analysis of the nature of science, usually resting on some aspect of the philosophy of science and its epistemology. Once these 'processes' are defined, they require to be represented in some test instrument. It tends to be assumed that such processes exist a priori, based on logical but not experimental grounds so that in the best case the next step is just a question of constructing an instrument with appropriate statistical properties, which allows such processes to be elicited and to be measured. That is, because the 'scientific processes' are taken as being pre-defined, the appropriate strategy seems to be to use the best psychometric resources available, so as to measure them reliably and validly. Looked in this way, the model of Figure 1.1 seems to offer a way of knowing how children perform on some 'scientific processes', often implemented by means of 'paper and pencil' tests, and therefore being in a position to judge to what extent certain educational objectives have been achieved.

In this linear model, the elements and relations between them can be taken as problematic. There are, for example, different analyses of the nature of science which can be (and have been) taken as starting point. There can be differences about the basic nature of the test instrument (paper and pencil, versus practical tasks). The linear nature of the model, however, represents these issues as needing to be resolved in a logical progression, solving each before proceeding to the next. In addition, the model has no way of starting from acceptable, interesting, or valuable pedagogical activities as they actually exist. It does not allow for deriving a rationale from existing practice.

1.2 A CRITICAL MODEL.

Figure 1.2 shows a schematic representation of the relationships between the elements which is intended to allow for a critical analysis of these relationships, and to put under scrutiny the main assumptions that are made when trying to define, elicit and assess 'scientific processes' as educational objectives.

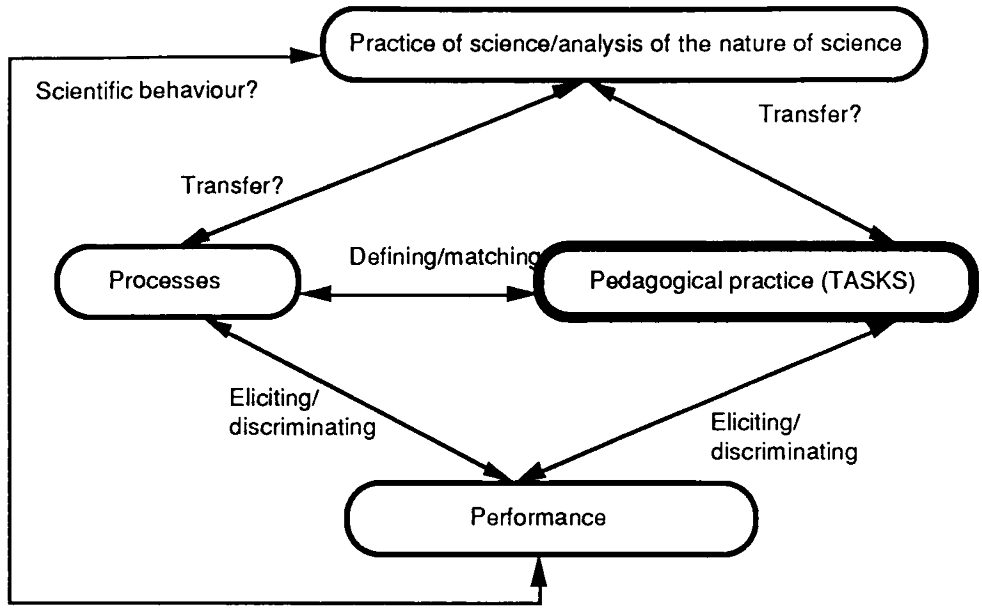


Figure 1.2: critical model of defining and eliciting children's 'scientific' behaviour.

The model shows essentially the same elements as those depicted in the linear model, with the addition of mutual interactions between them, the incorporation of scientific practice as source of reference and the setting of tasks in the context of pedagogical practice.

The elaboration of this viewpoint can be divided into two main branches, one theoretical and one experimental. The theoretical one includes the relationships of the philosophy and practice of science with 'processes', and with the choice of tasks in relation to pedagogical practice. That is, what teachers and scientists do is taken as an essential point of reference. Tasks given to children need to be justified from both points of view. The experimental, concerns the relationships between 'processes' and tasks, the relationships of both of these with children's actual performance, and the relationship of children's performance with scientific practice.

The discussion will begin with the theoretical issues, and continue with the experimental ones.

1.2.1 PROBLEMS OF TRANSFER FROM THE NATURE OF SCIENCE.

Figure 1.2 takes as problematic both the transfer from the nature and practice of science to defined 'processes', and to the nature of appropriate tasks. It does not take it as evident that educational activities follow directly from the nature of the subject matter.

1.2.1.1 Concerning processes.

There is a reason why 'processes' are currently so prominent in thought about science education for primary school children. It is the movement towards emphasizing the active role of individuals in constructing knowledge that has been restored in opposition to a passive view of learning. That is, active methods based on the methods of science have been implemented and preferred to those that emphasize learning the 'facts' of particular subject matters; leading to the stress on the so called 'scientific processes', and to their use as fundamental elements in structuring the curriculum, indeed lying behind the metaphor of children as scientists.

Having accepted that it is desirable to promote children's engagement in science activities, particularly those concerning the method or methods of science, two problems have to be faced. The first concerns the assumption that science can be characterized by its methods, and the second the problem of deciding from what analysis of science to draw the 'processes'.

The question whether science can be characterized by its methods (and whether they 'assure' true knowledge) is beyond the scope of this research needing an answer within the domain of philosophy of science or epistemology. The question of how valid it is to represent scientific activity by 'science processes' ('method') in the classroom is a different one. Science processes as part of learning about "How can we find out?" and "How do we know?", represent one line of development in science education (Ogborn, 1988a, 1988b); other lines being learning about "What are things like?" or "What are they made of?" (the ontological line), "How does it work?" (the causation line), "What does it mean?" (the communication line) and "What can we do ?" (the pragmatic line).

Assuming that 'scientific processes' can, to certain a extent, validly represent scientific activity in the classroom two problems arise:

1) whether it can be assumed that a simple and direct transfer from a logical reconstruction of scientific activity (philosophy of science: whether from context of justification or discovery) to an educational environment (science education) is possible and,

2) whether it is appropriate, as has happened, that 'scientific processes' can be identified by drawing mainly on one position, namely the inductivist-empiricist tradition.

The first problem rests on the prior assumption, that what is valid in one context (philosophy of science) is necessarily valid in a different one (science education). In other words, that the practice of and justifications given by professional scientists have the same nature in children's educational activities with a 'scientific' content inside the classroom, neglecting their differences in nature, purpose and constraints. To make explicit this distinction Martinand uses the term 'practice of reference' ('pratique de référence') to refer to a context (practice of professional scientists) different from the educational one (Martinand, 1983, 1986). A distinction has also to be made between the actual performance of scientists when engaged in a scientific problem and the logical reconstruction of their actions by philosophers in trying to address the problem of how scientific knowledge is achieved - that the context of discovery is not the same as the context of justification (Ziman, 1984). Similar distinction has to be made between the actual children's performance and the representations made to explain how children go from one level of understanding to a more complex one, as part of a different practice, the educational one.

To better understand the differences between a professional practice and a pedagogical or educational one, the three differences mentioned will be described further. The professional practice usually implies a specific training to become a member of a particular community in a given field of study; it is a way of living. Children at school are at best just trying to grasp the basics of science; they do not make their living by performing the tasks set in the laboratory or in the classroom. The purpose of the two practices is different. One is trying to generate reliable and valid knowledge which stands replication and criticism, and the other one is trying to gain some learning or 'insight' in relation to knowledge that has already been 'produced' elsewhere. The constraints are different as well, because in professional practice scientists are helped in their purpose of getting reliable knowledge by criticism from within the scientific community; something which makes the scientific enterprise not only a personal activity but also a social one. On the other hand children's work is judged individually, not as to whether it stands criticism or not; but by evaluating if they attained particular learning goals. That is, scientific activities for children are to be seen as contributing to their personal development; for scientists as contributing to the production of knowledge. This difference will inform the discussion of the empirical work undertaken in this study.

If the differences between both practices or contexts are accepted, 'scientific processes' can not just be transferred from one to the other. What is involved is what Chevellard has called a 'didactic transposition' ('transposition didactique') (Chevellard, 1985; Martinand, 1983). The didactic transposition implies that the original practice will suffer a process of 'decontextualization' ('décontextualisation') - because of changing context - and a process of 'recontextualization' ('recontextualisation') - because of adapting to a different context -, if an appropriate transfer is desired. It may well be that both practices could be 'similar', but analytically they are qualitatively different. Obviously, there remain the problems of deciding what the didactic transposition acts upon (what processes would be proper for young children), and of finding out experimentally how effectively and appropriate they are for children at a given age. This last is central to the present research.

The second problem, of whether the inductivist-empiricist tradition should be the main primary source of ideas about possible 'processes' in the educational context leads us to consider whether at least two more should be recognized: the rationalist and social constructivist traditions. It is certainly not my intention to solve this problem at the level of epistemology or philosophy of science, but I shall adopt the view that there is no such thing as '<u>the</u> scientific method'; but that there is a diversity of them.

Within the inductivist-empiricist tradition it was assumed that there was a unique method, the so called 'scientific method', that could whatever the content of the science, be applied in order to assure reliable and valid knowledge; a view that now seems to be history. Nevertheless, the effects of such a conceptualization reached the educational context in the form of assuming that processes are independent of content; allowing for structuring entire curriculums based on this idea (Science- A Process Approach (Gagné, 1965); Warwick Process Science (Screen, 1986). Because the content is seeing as being processed by a method which gives automatic assurances of reliability and validity, these processes appear to be a secure basis for a curriculum. The present study is designed to produce evidence of the relationship between process and content. Even if it were the case that in the context of science there was one valid scientific method, independent of content, such an assumption is not necessarily valid in the educational context: experimental evidence would be needed to support any claims that the transposed 'processes' also exist independent of content.

The question of where to draw processes from remains open and, as was said, two different traditions can be added. A more rationalistic approach would perhaps emphasize processes like conceptual change (Carey, 1985) or expectations, and a social constructivist one would offer processes like exchange of information (implied in Knorr-Cetina's 'discursive interaction' in the construction of knowledge (1981), in Ravetz's 'quality control' in science (1971) and Millar's 'negotiating knowledge' (1987)). These two examples clearly show that the source of processes is not limited to just one tradition, and that there is no reason why they should not be mixed in the same model within an educational context. Further, it can be admitted that valid processes can also be found, not in science, but in pedagogical practice. Classroom practice may have a rationality of its own, independent of that of any source discipline.

1.2.1.2 Concerning the eliciting tasks or devices.

In the linear model, the test or eliciting instrument derives its nature from something else - as discussed above. If the nature of a task is to be taken as problematic, this raises the possibility of looking to existing valued classroom practice as a source of tasks. To do this would be to rely on the intuitions of teachers as to what it is appropriate to ask children to do. Of course, this could not be the only criterion: the point of the alternative model is to allow such a possibility to exist, but then to require tasks devised in this way also to be looked at in relation to an analysis of science and its practice and in relation to possible 'processes'. But now tasks are allowed to suggest processes or to select aspects of science for attention. Further, as in the case of 'processes', there are several ways to interpret the essence of the scientific enterprise that can provide with inspiration for pedagogical tasks, such as an activity that discovers the laws of nature, as a historic reconstruction of knowledge, as an investigating process or as a problem solving activity for example; the study implements tasks as an investigating process plus a clear goal to be achieved. The arrows linking tasks to the nature of science and to processes go in both directions.

In the present work therefore, tasks will be proposed and investigated, which start at least in part from 'scientific' activities to be found in primary schools. Data about their acceptability to teachers will be taken as one criterion of appropriateness. Put fundamentally, this means that the design of the tasks places as much emphasis on the rationality of practice as it does on the rationality derived from some analysis of science.

This raises a problem. In the linear model, we require the test instrument to have adequate psychometric properties before it is considered functional. In the alternative model, the statistical properties of scores derived from tasks are one, but only one, of the kinds of evidence needed to develop a critical discussion of the meaning and value of 'processes' reflecting 'scientific activity' as realised in tasks which also have some classroom validity.

A further problem is that a task chosen as a valid classroom activity for children will not necessarily be well adapted to generating clear cut data about all aspects of processes which one may need to investigate. Indeed, this is the case in the present study, leading to the development of two forms of tasks, one of which produces clearer-cut data but which has less evident validity as a classroom activity.

1.2.2 PROBLEMS OF DEFINING AND MATCHING.

Even if the previous problems can be solved, there still remains the problem of representing processes and images of science in the tasks; with the inherent problems of defining and matching processes and tasks if the question of whether processes are in any sense independent of content is to be addressed.

Let us address the problems of defining processes and matching them to tasks, first. One way which has been used of going from processes to tasks is to develop a paper and pencil instrument, often with the intention of meeting some psychometrical criteria. In such a situation the definition of processes is constrained, basically, by the feasibility of representing physical situations by written and graphic means and, by children's ability to understand these. Such a procedure may well facilitate important assessment characteristics like reliability, but faces problems of validity (understood as the closeness between the nature of what is elicited by the instrument and the use of the ability in 'real' situations).

It is of course not necessary to use a paper and pencil instrument. But the direction of matching processes to tasks necessarily tends to produce relatively artificial and fragmented tasks, each piece matching a process but not necessarily integrated as a whole.

The alternative direction of defining and matching is from task to process - to choose a task which as a whole has the possibility of stimulating and eliciting a number of processes. This produces different problems: there is no guarantee that such a task will elicit all of the processes one may hope to investigate or assess and, some processes may be difficult to elicit because of the spontaneity and continuity required in the task. There is a price to pay for 'validity'!

In the present study, one form of task is of this second (alternative) kind. However, it is supported by a second form of task in which processes are represented more systematically in the design of the tasks. Because both share the same content it will be possible to bring out some evidence about the differences and connections between such forms of task.

1.2.3 PROBLEMS OF ELICITING AND DISCRIMINATING.

These problems, as shown in Figure 1.2, concern the relations between performance and tasks and performance and processes. Both involve children's actual performances. Having decided about where to draw processes and images of science from, having selected the appropriate processes and tasks (which also means deciding about their didactic transposition), there remains the question of comparing what is actually elicited against what was intended to be elicited.

In the linear model, where processes are pre-defined, the job of the test instrument is to realise them operationally. Of course, the attempt to construct the instrument may fail, in which case one might want to conclude that the processes have no real 'existence' in the way children actually function.

In the alternative model one can, as above, succeed or fail in getting evidence of previously defined processes. But in this model, the definition of 'processes' is taken as problematic, so that evidence from performance can be used to alter the definition of a 'process', or even to introduce a previously undefined process. So data on performance, on this model, is not just data to be reported, assessing children, but contributes to the critical discussion about the nature of 'scientific processes in children'.

1.2.3.1 Processes and performance.

The basic question here is to know to what extent children's behaviour resembles the intended processes to be elicited. This question becomes increasingly important if the problem of content independence is addressed, because the idea of the intended processes rests on a supposed similarity exhibited across different contents.

What can be elicited also depends on the form of the task or instrument.

In the extreme case of a multiple choice instrument, the chosen answer from those offered does not necessarily reflect how children think. Then the problem of how far children's answers are from those intended to be elicited seems to dissolve. It is reduced to find whether children can or can not identify the best required answer as defined before hand; it is in gross terms a matching between the adults bias towards what is considered as important and the child's intention to discover what the adult wants as an answer. The same applies when the same item is set across different contents. If on the other hand, the task permits a free response, whether in writing or in actions, the problems are different. A task requiring writing, although offering children the opportunity to express themselves and putting the teacher or the researcher in a position to judge children's responses in the light of how near they are or not to what was expected, poses the difficulty of expressing thoughts in writing, and the restrictions imposed by the use of the language. But the important thing in common with tasks requiring actions as a response, they allow us to compare what is expected to what is actually elicited. As a consequence, they can offer evidence to decide whether professional performance (coming from scientists) differs qualitatively from children's performance. The problem becomes more interesting when elicitation of the 'same' process by this means, involves different contents; allowing it to be seen whether children give the same kind of answers or produce the same kind of actions.

From the point of view of Figure 1.1, the question whether a process exists or not is to be decided essentially on statistical grounds: do performances correlate or not? Psychometric theory requires a certain degree of variability in children's responses in order to be able to detect correlations, and so as to maximize individual differences for assessment purposes. Therefore, in statistical terms, very easy or very difficult 'processes' can not exist in this sense, regardless of how consistent they might be across contents. A curious consequence is that a 'process' which has this statistical 'existence' at a certain level of development (when it shows an appropriate degree of variability), can cease to 'exist' once everyone achieves it. Following this line of reasoning, somebody who has the ability to speak Spanish would not be detected as having this ability among the Spanish population, but it would certainly be among an Anglo-Saxon population. Results from the study will confront this issue, because certain 'processes' turned out to be very easy, and others very hard.

The statistical study of 'processes' also requires performances to be scored on some scale. This should not be taken as unproblematic, and is necessarily a difficulty when performances are free responses. The study will examine what is involved in arriving at scores for qualitatively different kinds or levels of performance. The relation of scores to what can reasonably be expected of children in a given task will be an issue to be discussed.

1.2.3.2 Tasks and performance.

Eliciting 'scientific' behaviour from activities such as doing an investigation, has its own problems. If children are working spontaneously on their own, it is difficult to ask directly some questions. Some of their 'scientific' behaviour can perhaps be inferred from what they do, but there still remains the problem of what is to be considered as 'scientific' behaviour. It still has to be seen whether professional and educational performances are qualitatively different or not.

1.2.4 CHILDREN'S PERFORMANCE AND THE NATURE OF SCIENCE.

This section concerns the last link in Figure 1.2, between what children do, and what scientists do. The discussion is important because of its significance for critical thought about 'children as scientists' used frequently in approaches to primary science.

The linear model assumes that what is being elicited is indeed scientific behaviour, and the question is just quantitative: how much of it is there? But the deeper question is to what extent children's behaviour in any way resembles a professional one. Is it the case that children have their own distinct and different ways to solve problems and to investigate 'scientific' problems? To address this problem, it is essential to treat it as problematic how far children's performance matches, or otherwise, some expected kinds of behaviour. This the study is designed to permit.

1.3 SUMMARY.

The chapter has described and analyzed the problems of defining and eliciting 'scientific processes in children', starting with sources of possible processes and ending with evaluating the nature of children's responses; using the opportunity to mention in each case where the study hopes to contribute. It is clear that the study has different levels of concern at different points of the description and analysis. Here it is intended to summarise the problems with which the study is concerned, and to state where they are going to be discussed in more detail.

- Problems of transferring:

As mentioned, there are two problems: whether science can be characterized by its methods in educational practice and from where to draw processes.

It should be stressed, first, that this study is concerned with procedural knowledge (processes); although some aspects related to conceptual knowledge are not ruled out as a consequence of addressing the problem of whether performance is independent of content or not.

The question whether 'scientific' activities can validly be thought of as just method independent of content is not going to be discussed here, but it is assumed that such a representation is not complete if conceptual knowledge is not taken into consideration as well. The literature review will discuss attempts to represent science, in an educational context, uniquely by its methods, as well as the more popular sources from which processes have been drawn, and will discuss the problem of identifying as actual processes a logical reconstruction of scientific activity.

In relation to the use of tasks based in part on classroom practice instead of tests instruments relying on pre-defined processes the literature review will describe attempts to implement both modes. The attempt in this research to build feasible practical activities as tasks to elicit processes will be dealt in Chapter 3, and the main results are given in Chapter 6. The questions of what 'scientific processes' were elicited and how they relate to success on the tasks can be found in Chapters 7 and 9 respectively.

Given the decision to construct tasks taking classroom practice into account it was important to show how far they agreed with such practice. Primary science teachers were asked about their expectations of children's performance on the tasks, and on their ratings of scientific processes as educational objectives. Results can be found in Chapter 3.

- Problems of defining and matching:

The problem of defining processes starts when a definition of what is intended to be elicited is given and continues through the different subtasks designed to elicit the same putative process and the decisions to combine scores as part of the same dimension. This is discussed in Chapter 5, in which the construction of scores is both described and taken as problematic.

The question of matching processes for different contents is addressed in Chapters 3 and 5 where the problems in constructing tasks and defining scores for the same intended process for different contents are discussed.

- Problems of eliciting and discriminating:

The problems of eliciting and discriminating 'scientific' behaviours with similar nature and different degrees of variability to be regarded as processes are discussed in Chapter 7. This chapter also discusses to what extent it is possible to elicit spontaneous processes coming directly from the tasks.

- *The problem of the nature of children's 'scientific' behaviour:*

The question of how far children's behaviour resembles behaviour of scientists, or of a logical reconstruction of their behaviour, is discussed in Chapter 8.

Chapter 2.

MAIN ISSUES IN RELATION TO 'SCIENCE PROCESSES' AS ELEMENTS OF THE ELEMENTARY SCIENCE CURRICULUM.

The inclusion of scientific processes, giving them various meanings, as educational objectives in elementary schools has a long tradition. Their appearance in elementary curriculum developments as the main learning targets has been intermittent, depending on the emphasis given to learning the concepts, content or facts of science.

The importance given to science processes and practical activities by science curriculum development projects, assessment schemes, educational researchers and governmental agencies, will first be reviewed in general terms.

Secondly, the role of science processes and the importance of practical activities, in some of the most important curriculum projects for elementary schools (with preference given to those particularly addressed to primary schools), and for assessment schemes, will be reviewed in more detail.

Thirdly, I will discuss what have been identified as some of the main issues in relation to defining and eliciting science processes. The intention is to focus on them individually, but without losing sight of them as part of a system that goes from selecting and 'transferring' 'scientific' processes to analyzing the nature of children's performance on these processes. Thus, the structure of the previous chapter will be maintained in this section of the review.

It is widely recognized that 'scientific processes', and 'process skills' have been given many different meanings, and that this fact makes it difficult to deal with them (Doran, 1978a); Donnelly & Gott, 1985; Millar & Driver, 1987; Jenkins, 1989; Fairbrother, 1989). In this review they will be referred to broadly as science processes, being understood variously as those activities undertaken by scientists in their work as professionals, as those elements identified as important in getting reliable and valid knowledge by certain philosophies of science, or as those cognitive abilities of kinds recognized as relevant for the understanding of science.

2.1 IMPORTANCE GIVEN TO SCIENCE PROCESSES AND PRACTICAL ACTIVITIES.

For a long time, teaching and learning the processes used by professionals of science has been one of the major objectives of science education in schools.

Since the middle of the 19th century, science educators have argued that the processes of science should be taught as part of the school curriculum; in this country Thomas H. Huxley among others. That period is described in Laytons's book 'Science for the people', quoted by Finley (1983):

> "The unique characteristic of science as branch of learning was the method by which knowledge was acquired [and that] the inductive aspects of science activity, rather than the conclusions were of most significance from an educational point of view. Science was [to be] studied in the school not for its informational benefits but because it trained the power of observation and reasoning". [Layton, 1973, p. 172]

In 1867 the British Association for the Advancement of Science (BAAS) published a report "On the Best Means of Promoting Scientific Education in Schools", and as Jenkins (1989) points out, its rationale was underpinned by a sharp distinction between 'literary acquaintance' with scientific facts and the knowledge of methods that may be gained by studying the facts at first hand.

The method of science, or better the methods of science, has a close relationship with science teaching in the laboratory. The United Kingdom has a long tradition in this sense, as Jenkins establishes: secondary school labs date from 1877; their role, to back up the intention to acquire the 'scientific habit of mind', that is the mental training and intellectual discipline characteristic of the practising scientist. Much of the influence on practical laboratory teaching in secondary schools is due to H. E. Armstrong and his heuristic approach to science (based upon 'the scientific method') in the late 19th and early 20th centuries. But such an approach to science, which implied a correspondence between the knowledge obtained from using the scientific method and the real world, could not explain (as Jenkins points out) developments such as the quantum theory and others, in which the concepts and imagery of science moved far from common sense. This problem was addressed by the BAAS, stating in 1917 that 'the scientific method is an abstraction which does not exist apart from its concrete embodiments' (Jenkins, 1989). This shift from 'process' to 'content', supported by the idea that students are to be introduced to the intellectual constructions of science, was encouraged 'by the findings of experimental psychologists whose investigations undermined the transfer of training upon which heurism rested' (Jenkins, 1989).

By the end of the 1950's, the demand for more qualified scientists and the space race brought 'processes' back again to elementary curriculum developments: the Nuffield Science Teaching Project in the U.K. and the 'Science - A process approach' (SAPA) in the U.S.A.. In this climate, 'training in "scientific method" as a curriculum objective was strongly reasserted' (Jenkins, 1989). In his view science curriculum developments in many parts of the world at that time can be characterized by their emphasis on scientific procedures and attitudes ('investigative', 'open ended' or 'discovery' learning). Jenkins (1989) agrees with Richmond and Quraishi in characterizing the early years of the 1960's as 'neo heurism' (referring to Armstrong's heuristic approach to science).

Supported by the Nuffield Foundation, the Junior Science Teaching Project started in 1964. It emphasized practical experience (exploring and experimenting) guided by a Piagetian perspective. In 1967 the Schools Council and the Nuffield Foundation sponsored the 'Science 5-13' project; it shared the concern for matching children's cognitive development and educational demands, as well as the emphasis on practical activities of the Nuffield Junior Science Project (Harlen, 1975).

In 1965 Robert Gagné presented the psychological basis for another curriculum project, SAPA, to the American Association for the Advancement of Science (Commission on Science Education), based on the the 'processes' of science. He argued that prerequisite scientific concepts and principles are to be obtained only through the operation of science processes such as observing, classifying, inferring, etc. (Gagné, 1965); an approach which appears to be based on an inductivist and empiricist view of science (Finley, 1983). This project is very well characterized by Karplus & Thier (1967) when they say that it 'stresses the child's practice with the processes and uses the phenomena only as vehicles and the concepts as tools'. The content of Gagné's paper has had a substantial influence on curriculum development, instruction and research in science education. The influence of Gagné's ideas is complemented by Bloom's attempt to give education a rationale through his behavioural objectives; a scheme which has proved very attractive for objective assessment and for guiding pedagogical practice. The influence of such attempts reached this side of the Atlantic in guiding the attempts of 'Science 5-13' to establish clear educational objectives which matched with children's cognitive development, although drawing more from Piaget than from Gagné for its psychological basis.

Other projects that appeared in the 1960's in the U.S.A. were 'Elementary Science Study' (ESS) developed by Educational Services Incorporated, and 'Science Curriculum Improvement Study' (SCIS) established by Karplus, professor of theoretical physics at the University of California in Berkeley. The first, backed by Bruner, 'stresses the child's involvement with the phenomena and is confident that he will thereby gain practice with the processes and achieve understanding of valuable concepts even though these are not made explicit' (Karplus & Thier, 1967). The second, SCIS, is described by Karplus and Thier (1967) as a project that 'stresses the concepts and phenomena, with process learnings an implicit by-product of the children's experimentation, discussion and analyses'. It too is supported by a Piagetian framework.

There is another project, relevant to incorporating science processes in the curriculum, which has recently appeared (1986) in the U.K.: the 'Warwick Process Science Project'. As defined by Screen (1986), it is 'a process-led (*rather than a knowledge-led*) science course' (my italics). Its similarity with SAPA, in relation to the processes advocated, is remarkable.

At the end of this list of curriculum projects comes a recent one (1990) called 'Science Processes and Concept Exploration Project' (SPACE). It is a classroom-based project that intends to map the ideas which primary school children have in particular concept areas, and to establish the possibility of children modifying their ideas as a result of relevant experiences (SPACE, 1990). Eight concept areas have been studied so far, with no reports on science processes yet.

The Assessment of Performance Unit (APU) has surveyed 11 year olds (and others) since 1981, and has 'processes' as one of the 'facets of science performance' to be assessed in its scheme, as well as concepts and attitudes (DES/APU, 1988). Although it is not a curriculum project, its inclusion here is justified on the grounds of its influence on science teaching through the publication of its monitoring results and of the kind of tasks used to assess processes on a national scale. Particularly interesting is the attempt to assess performance on practical investigations.

Another model of assessment developed in the U.K. is the 'Techniques for Assessment of Practical Skills in Foundation Science' (TAPS). This project has developed assessment instruments for evaluating practical work that emphasize practical skills used in the laboratory, but which are not restricted to manipulative dexterity (Bryce & Robertson, 1983; 1985). The way they conceptualize practical skills 'in accord with the way in which teachers think and act in the the school laboratory' is interesting; it looks like a more pragmatic approach, rather distant from the philosophical ideas which have influenced others.

In the same assessment perspective, the Dutch Institute for Educational Measurement (CITO) is said (Hellingman, 1982) to be moving from paper-and-pencil-tests towards more or less open-ended pupil achievements, including those in practical work; thus facing the problem inherent in open-ended practical investigations, that 'of blurring the distinctions between practical and theoretical work' (Hellingman, 1982).

Governmental agencies, such as the Department of Education and Science (DES), have also expressed their concern about science processes. As is made explicit in 'Science 5-16: A statement of policy' (DES, 1985): 'the essential characteristic of education in science is that it introduces pupils to the *methods* of science' (my italics).

Recently, some science researchers have published their views about science processes: Finley's 'Science Processes' (1983), Adey & Harlen's 'A Piagetian Analysis of Process Skill Test Items' (1986) and Millar & Driver's 'Beyond Processes' (1987). Finley concentrates on the epistemological foundation of Gagné's conception of science processes, deriving some possible consequences for learning science, such as that 'there is a clear danger that students will be presented with an inaccurate and inadequate view of science processes' (Finley, 1983). He criticizes the empiricist and inductivist view of science held by Gagné, and argues in favour of a view that is 'conceptually driven' - seen as more hypothetic-deductive rather than inductive. By doing so, he attacks at an epistemological level the notion of a single 'scientific method' (that involves certain processes) used by scientists in different disciplines and topics, and points out the unlikelihood that 'there will be content-free intellectual skills that are generalizable across multiple enquiries'.

Adey & Harlen analyzed test items (developed by the 'Assessment of Performance in Science' (APS), used to survey attainment on certain science process skills in a representative sample of 11 years old in Britain, to determine their level of cognitive demand in Piagetian terms. They used three of the APU item categories (process skills: use of symbolic representation, interpretation and application and planning of investigations) as objects of the analysis, and the 'Curriculum Analysis Taxonomy' (CAT) (devised by the 'Concepts in Secondary Science and Mathematics Project' (CSMS), and described in Shayer

& Adey, 1983) as the instrument to determine the level of test item cognitive demand. Their main claim is that strong evidence was found of the predictability of the difficulty of item testing process skills using the CAT, based on Piagetian ideas of stage development. Another claim is that 'there is support for the notion of a hierarchy of process skills'; which may remind us of Gagné's (1963, 1965) hierarchy of 'basic' types of process skill and of the other higher level 'integrated' type of process skill (a hierarchy for which Gagné does not offer empirical evidence; see below). But what is interesting is the fact that the CAT uses a very different set of 'processes' in its approach, based on Piagetian work. Their 'processes' are divided into two groups: a) psychological characteristics of children's thinking relevant to the understanding of science, and b) intellectual elements or schemas specific to different types of science activity (Shayer & Adey, 1981). Thus, Gagné's list of 'processes' is but one of several that can be set; his list has five 'integrated' processes, compared to CAT's list of nine schemas and rather little overlapping (see below). This points out that the selection of processes to attend to is problematic.

Finally, Millar & Driver (1987), correctly in my view, address the question of 'science processes' critically. Three levels of discussing science processes are taken on board: epistemological, psychological and pedagogical. The first level (see also Section 1.2.1), is not within the scope of this study although the possible implications of philosophical views adopted in relation to science teaching and learning are recognized. In relation to the second, they make two points: a) what children learn from interaction with phenomena depends not only on what is abstracted from the situation but also from the 'mental constructions' that they bring to it and, b) the existence of science processes, or as they call them 'process skills', are not content and context independent. Both points require empirical evidence to be substantiated or refuted. There is an extensive literature showing the evidence collected in relation to what children bring to situations in the form of 'misconceptions', 'alternative conceptions', 'alternative frameworks' etc. in different domains of science. It still remains to be explained what are the mechanisms of conceptual change, and whether certain cognitive abilities are involved in such change. The question of the existence of 'science processes' as independent of content is a basic one, and needs, as stated above, an empirical basis. Essential questions are, "How are they defined?", "How are they elicited?" and "How is the matching of the same intended process for different domains resolved?" The third, pedagogical level discussed by Millar and Driver, is in my opinion the most important one, if 'science processes' are to be anchored in pedagogical practice. Millar & Driver are sceptical of the idea of cognitive skills being taught and transferred from the context where they were learnt to a different one, and challenge the meaning of progression in learning or performing such skills. Certainly teaching cognitive abilities for their own sake, presents problems of transfer and progression. More important is the question, "What for?". In my opinion, such abilities only make sense when applied in a more or less open situation containing some kind of problem or requiring some kind of practical investigation, that is in a task to be used in the classroom.

It can be concluded at this stage of the review that:

i) great importance has been attributed to 'science processes' as educational objectives, as targets of assessment, as structuring elements of elementary curricula, and as matters of educational policy, but with too much variety for such a movement to be integrated or coherent;

ii) science processes can have different sources of inspiration and can be discussed at different levels of analysis (epistemological, psychological and pedagogical);

iii) the use of practical activities in learning science in elementary school is recognized as of prime importance;

iv) after so much advocacy of science processes, we are still unsure how such processes come to exist. And, if they do, what is their pedagogical meaning?

v) there seems to be a relationship between science processes as educational objectives, as practical activities implemented in the classroom and for assessment, but what this relationship is is not clearly established;

vi) there seem to be two main views of teaching science in elementary education: one puts emphasis on the 'results' or 'products' of scientific activity, that is on knowledge - complicated by the fact that children bring their 'own knowledge' with them -, and the other focuses on the 'means of getting knowledge', that is on ways to 'shape' children's minds - complicated by evidence of commonsense modes of reasoning in children. The big challenge seems to be to understand how they are connected.

These introductory and general conclusions will be made more specific in the following sections of this chapter.

2.2 SCIENTIFIC PROCESSES AND PRACTICAL ACTIVITIES IN SOME CURRICULUM PROJECTS AND ASSESSMENT SCHEMES.

Science processes have been an important element in educationalists' attempts to make children acquire certain behaviours deployed by scientists, by introducing certain kinds of task into the curricula. Thanks to Bloom's intention to provide a rationale for education, behavioural objectives are now part of the educational culture. Accordingly, assessment of such objectives is often seen as an essential part of a curriculum. But, there seems to be a gap between the curriculum objectives, the assessment of such objectives, and the tasks carried out in the classroom. The emphasis has been either on the objectives or on the assessment, and not on what it is possible to do or to elicit from activities performed in the classroom; where the individual's construction of school knowledge is supposed to occur.

A closer review of the role of processes, of their origin and construction, of the kind of tasks encouraged, and of the way curriculum materials are evaluated in certain curriculum development projects and schemes of assessment, will now be given, focusing on the gap already indicated.

2.2.1 CURRICULUM DEVELOPMENT PROJECTS.

Nuffield Junior Science (NJS):

The purpose of this project was to develop teaching materials, which took the form of a series of published books addressed to primary schools teachers (mainly the lower grades) offering them information and guidance.

It insists on practical experience (problem solving) as a way of learning, guided by Piagetian findings such as that: children are able to perform a problem in logic practically, but usually incapable of doing the same with simple verbal propositions. Similarly, the formation of concepts is 'through practical experience' and develops at the child's own pace. Two kinds of experience are considered: exploring and experimenting. In their philosophy of learning science certain 'processes' appear (though no specific framework to which they can be attached is mentioned): 'isolating a problem', 'planning an experiment', 'observing critically', 'discussing', 'representing ideas' and 'communicating', among others. The approach, which is 'very much open-ended, free-ranging enquiry in science makes syllabuses in primary schools impossible and, indeed, demands a fresh approach to the drawing up of records of work for individual children' (Nuffield Junior Science Project, 1970).

Science 5-13:

This project focused on children from primary to early secondary school. It wanted to help teachers in selecting experiences and activities appropriate to children's stages of development - to gain firsthand experience through discovery methods. It wanted to establish clear objectives to guide pedagogical practice. A list of scientific concepts, specific knowledge, 'processes' and attitudes was drafted indicating what might be expected at different Piagetian stages of development. Following the Bloomian idea and structure of objectives, specific objectives were logically derived from more general ones for example: Ø.40 'posing questions and devising experiments or investigations to answer them'; 1.42 'ability to make comparisons in terms of one variable'; 2.43 'appreciation of the need to control variables and use control in investigations'; 3.42 'ability to design experiments with effective controls for testing hypothesis'. Other 'processes' being: 'observing', 'communicating', 'appreciating patterns and relationships' and 'interpreting findings critically' (without any claim attaching them to a specific method in science) (Schools Council, 1972). Following the first draft of the objectives, they were tried in schools, discussed with teachers, modified and tried again. The project relied heavily on Piagetian findings to establish the 'specific aims', showing now in retrospective the problems of knowing with any certainty what could be expected in relation to 'processes' in different contents.

The project developed some assessment instruments to gather information about changes in children's achievement in relation to the established objectives. They used a 'non-conventional' test. As Harlen (1975) says the most obvious way would be to present children with a 'real situation' that required a particular 'process' or concept in order to deal with the situation successfully; but that was regarded as impracticable, as it would require individual administration. To avoid too much dependence on verbal ability and some other disadvantages of the paper and pencil form of test, a film of an experiment was used. This way of representing the problem was considered clearer than a verbal description (children's answers were in the form of ticking one of the alternative responses offered). The use of the problems or situations described in the units was avoided, looking in this way for application of something already learned in one content or context to a different one (transfer). No systematic comparisons within individuals in different contents are reported, but only overall effects of the trial work. The value of the test results in revising the unit, as Harlen (1975) admits, 'was not very great': they 'had been useful in indicating overall achievement of stated objectives', 'but they could not suggest why there had been no progress or indicate ways in which these sections might be changed'.

Science - A Process Approach (SAPA):

This project set out to develop an elementary science curriculum, including a model for curriculum evaluation. The materials developed include learning units (with their objectives, activities and materials), their relationships with other units ('processes'), group and individual competency measures (tasks designed to measure levels of attainment) with corresponding acceptable behaviours (AAA, 1967), and the Science Process Instrument (a longitudinal test with diagnostic purposes, which remained in an experimental phase because of lack of funds). Processes are defined 'as those activities common to scientists of all scientific descriptions when they are practicing science' (AAAS, 1968). The evaluation of the curriculum is based on the assessment of instructional aims, expressed in Bloomian terms, as objective behaviours; these are organized as learning hierarchies. To establish these, the project stated the formal requirements for the designed tasks (organized around 'science skills') and then derived subsets of behaviours formally required to do the exercise, going from the most complex to the simplest (exactly as 'Science 5-13' did, but with different psychological basis). The project described how this was done as follows:

'The terminal behaviors have been identified from the set of behaviors for a given process. For each of these terminal behaviors, this question was asked-- what should the learner be able to do before it is highly probable that he will be able to acquire this behavior? This led to the identification of one or more subordinate behaviors. Repeating the question for each subordinate behaviour led to further identifications until all the behaviors associated with a given process were ordered into a dependency sequence' (Gagné, 1962, 1965; AAAS, 1968).

These requirements, thus, are decided a priori, taking the set of 'skill processes' from what has been identified as an 'inductivist' and 'empiricist' view of science (Finley, 1983; Millar & Driver, 1987). Eight 'basic' processes were defined ('observing', 'classifying', 'measuring', 'communicating', 'quantifying', 'organizing through space and time', 'inferring' and 'predicting'), as well as five 'integrated' processes ('making operational definitions', 'controlling variables 'formulating testable hypotheses', 'carrying out experiments' and 'interpreting data from experiments') (Gagné, 1965).

An evaluation model was developed to assess the curriculum materials. Tasks (individual and group competency measures) were developed to assess the behavioural objectives established for each process. The individual competency measures consists of two parts: a set of instructions which describes the assessment situation, specifies the tasks required of each child - usually requiring the use of apparatus - and describes the acceptable responses, and a standard response sheet on which to record results (AAAS, 1967). An experimental attempt was made to validate the learning hierarchies within each process (AAAS, 1968), but no attempts are reported to establish the relationships between processes, nor on their consistency within individuals in different contexts. Learning hierarchies have been severely critized for the way such hierarchies have been designed as logical pre-requisites rather than as a psychological sequence (Phillips & Kelly, 1975; Posner & Strike, 1976).

Science Curriculum Improvement Study (SCIS):

The idea of this project was that 'elementary school science had not only to be simplified but organized on a drastically different basis from the usual logical subject matter presentations to which the university scientist is accustomed' (Karplus & Thier, 1967). Karplus and Atkin (co-director of Illinois Elementary School Science Project) formulated a theory about guided discovery in which concepts are developed by being applied to new experience and not made as a result of an inductive leap. They pointed out the role played by preconceptions or 'mental set' of the observer. They state that such preconceptions determine which generalization a child can attain from his experiences. Karplus & Thier (1967) parallel scientific and intellectual development by saying that, 'the present content of science consists of concepts and relationships that mankind has abstracted from the observation of natural phenomena over the centuries' whilst during the elementary school years children 'accumulate experiences, and their thinking undergoes a gradual transition from the concrete to the abstract' (the Piagetian influence is acknowledged). Both the Piagetian influence and their conception of science seem responsible for the strong influence of concepts in the design of their learning units. The central objective of such materials was 'to lead children to approach the observation and analysis of natural phenomena by thinking in terms of interacting objects or components (*like a physicist*). At the same time, the activities in the classroom are intended to give children experience with a wide variety of natural phenomena, to develop many manipulative skills in the carrying out of experiments, and to furnish opportunities for recording the results of observation' (my italics) (Karplus & Thier, 1967). The most difficult decisions on each unit, they say, were about things that cannot be discovered by experiment: 'the man-made constructs'. To overcome such difficulties, it is said, one would have to have in mind 'what constructs are already available to the pupils and what constructs must be introduced to enable the pupils to make the discoveries potentially derivable from the experimental observations' (Karplus & Thier, 1967); in marked contrast with SAPA's view of science and children's intellectual features, where methods are far more important than concepts.

Studies were made of what children have learned from a teaching units. In this way they tried to find relationships between learning experiences and children's behavioural changes. They recognize the complexity of the objectives to be evaluated: 'a combination of understanding concepts, initiative in recognizing and attacking problems, and mastery of experimental techniques' (Karplus & Thier, 1967). They are aware of the implications for evaluating such complex objectives, stating that their evaluation procedures 'are based on confronting the student with a set of experimental problems for whose solution he has to apply the concepts and techniques developed in the unit. The student also has to explain why his procedure did lead to a solution. Some of the problems are open-ended; some have a single well-defined answer' (Karplus & Thier, 1967). The attainment of these objectives is seen as not directly observable, so as 'one must, therefore, decide what observable behaviors on the children's part will be accepted as evidence that the objectives have been reached' (Karplus & Thier, 1967). In relation to this difficulty, two problems seem to be present. Firstly, they miss the opportunities to elicit children's ideas, which they agree to be important. An example is when they admit they had to invent the concept of concentration for children (through a concrete operational definition: 'the darkness of the colored solution and the amount of residue left when the water in a solution is evaporated'). Secondly, the problem of establishing clear cuts for levels of attainment is not pointed out, despite the opportunity to do so. This can be judged from the kind of categories used to report children's answers in the case of 'separating a mixture of solid substances'. They report them as: 'no report', 'mechanical separation', 'mixed with water *only* ', 'used tea bag to immerse sample in water', 'mixed with water and filtered', 'mixed with water and evaporated the clear liquid', 'mixed with water, filtered, and evaporated the filtrate'; but it is not reported which ones were classified as reaching the objective or not or if different levels of attainment were considered, they only give group results for different grades and the comparison group that did not followed the teaching unit.

Warwick Process Science (WPS):

This project, in the words of Screen (1986), 'is a guide for teachers to *process work in science* and it provides teachers with ideas, materials and detailed instructions for a process-led (rather than a knowledge-led) science course, developing from the 11+ stage through to 16' (my italics). Processes, in this project, as in SAPA, are structuring elements of the curriculum; learning units are organized around each process. They are conceived as 'the sequence of events which are engaged when researchers take part in a scientific investigation', although they also consider other processes relevant for problem solving (Screen, 1986). The first set is: 'observing', 'inferring', classifying', 'predicting', 'controlling variables' and 'hypothesizing', forming a hierarchy of increasing intellectual activity and independence (although no experimental basis to substantiate this assertion is given). The others are 'planning', 'performing', 'interpreting' and 'communicating'.

Concepts in Secondary Mathematics and Science Programme (CSMS):

This programme, already mentioned for its Curriculum Analysis Taxonomy (CAT), although addressed to secondary children shows work on 'processes' from a Piagetian perspective. It attempted to 'investigate children's difficulties in science, by applying a Piagetian model of cognitive development to the practice of science teaching' (Shayer & Adey, 1981). In this sense they did not develop teaching materials as such. What they did was to develop some 'class tasks' ('Science Reasoning Tasks') in order to assess children's stages of development. This was used to assess the cognitive demand of curriculum objectives, allowing 'the selection of objectives and activities suitable for a given group of pupils, and also to rank levels of attainment within a topic' (Shayer & Adey, 1981). Based on the possibility of assessing children's stages of development, they tried to develop further the possibilities of curriculum analysis. Thus, the Piagetian schemas that underlay children's thinking were used as a basis for constructing a descriptive 'taxonomy'. Such a taxonomy classifies objectives into groups according to the schema or reasoning patterns employed, and the stage of cognitive development of which they are characteristic. To fulfil the first criterion of classification, Shayer and Adey took a number of different aspects of thinking from the Piagetian work, to cover a wide range of science activities . They offer two complementary categories for this dimension of the taxonomy: one which deals with six 'psychological characteristics of children's thinking' at each five stages of development, which 'is alternative and complementary to' the second that 'concentrates on the intellectual elements or schemas specific to different types of science activity' (Shayer & Adey, 1981). The first category contains: 'interest and investigation style', 'reasons for events', 'relationships', 'use of models', 'type of categorization', and 'depth of interpretation' (of descriptive passages). The second includes: 'conservation', 'proportionality', 'equilibria of systems', 'mathematical operations', 'control of variables', 'exclusion of variables', 'probabilistic thinking', 'correlational thinking', and 'measurement skills'. For each intersection of these categories and the cognitive stages of development, a description is given of what might be expected from children's performance. Although such descriptions have empirical basis (in Piagetian work), to my knowledge all categories have not been represented in different contents and children's behaviours elicited; what is reported (Shayer & Adey, 1981) is their attempt to validate Piaget's formal operations construction (basically that there is unity between schemas), by testing consistency between schemas in different contents (only three schemas are present in two tasks out of five, with five just in one task). What is interesting to note, however, is the possibility of representing 'science process' on a basis different from an inductivist-empiricist view of science; something also done by Tobin (1984) and Garnett et al. (1985). In fact Padilla et al. (1983) show an overall significant relationship between integrated science processes and formal thinking abilities. Yeany et al. (1986) report hierarchical relationships among modes of reasoning (Piagetian) and integrated processes skills (SAPA), using as Padilla et al., a multiple choice format for the instruments; such claims are also made by Adey & Harlen (1986).

Permantapan Kerja Guru (PKG) - Improvement of Science Teaching in the Secondary General Schools Project:

This project made 'learning by doing' its major objective and developed, and classroom tested 350 experiments for junior secondary science and senior secondary physics, biology and chemistry curricula. The intention was to create lasting experiences for students in applying science knowledge and in learning science skills.

An external evaluation team was appointed to evaluate the PKG project. This team, led by Professor J. Eggleston, developed an instrument ('Group Practical Tasks') in its attempt to 'observe and measure the abilities of students to collaborate, in their usual lab-groups, to solve new practical problems in a scientific way' (Aylward & Eggleston, 1986). In total ten practical tasks were developed, and a scoring system was constructed by studying video tapes of trial groups attempts at the tasks. Supported by repeated viewings of the tapes, the team constructed 'low inference check lists of observed items of behaviour'; allowing a qualitative and quantitative comparison of performance between different groups tackling the same tasks. The potential pedagogic value of such tasks was recognized by the team when looking at the videos, but their main purpose was to be used as diagnostic devices.

The tasks require the students 'to follow instructions, plan and organize their work, to observe, to make inferences from data recorded from their observations and to report their results and answers to the problems' (Aylward & Eggleston, 1986). The team examined each task focusing on the processes demanded for its 'definition', 'planning', 'execution', and 'interpretation'. It reports high levels of internal consistency for each of the practical tasks (Eggleston, 1984; Aylward & Eggleston, 1986). Although it is clear that the tasks had no intention to be used as tests and that they are group tasks, they might offer (with some changes) the opportunity for individual assessment and a chance to analyze the nature of the 'processes' elicited in each task (reflected in the check lists). the fact that tasks were devised for different contents makes it possible to see whether more general statements can be made about children's performance at this schooling level; particularly if these tasks are recognized as having a pedagogical potential.

Science Processes and Concept Exploration (SPACE):

This is a 'classroom-based research project', 'based on the view that children develop their ideas through the experiences they have', and has two aims: a)'to establish (*through an elicitation phase*) what specific ideas children have developed and what experiences might have led children to hold these views' and, b) 'to see whether, within a normal classroom environment, it is possible to encourage a change in the ideas in a direction in which will help children develop a more "scientific" understanding of the topic (the intervention phase)' (my italics)(SPACE, 1990). It is worth noticing that this is the first curriculum development project (of those reviewed here) that includes an elicitation phase in order to know what children's ideas are; instead of deriving then logically from some a priori conceptualization. The nearest mention of processes are those strategies included in the 'intervention phase': 'encouraging children to test their ideas' and 'testing the "right" idea alongside the children's own ideas' (in which 'children were given activities which involved solving a problem'), but no explicit hint is made of the project's view of the nature of science processes (SPACE, 1990).

2.2.2 ASSESSMENT SCHEMES.

Assessment of Performance Unit (APU):

The Science monitoring programme is one of the groups, within the Assessment of Performance Unit (APU) that belongs to the Department of Education and Science (DES), in charge of surveying performance in different educational areas. Its task was to carry out national sample surveys of pupils' performance levels in science. The review of this programme will contain two parts, one related to the work of the programme as a whole, and the other in relation to the assessment of 11 year olds.

One of the major challenges undertaken by the Monitoring Team was to construct an assessment framework. Additionally, it inherited an *across-the-curriculum* view of teaching and learning from the Science Working Group, having being decided that a traditional heavily content-based assessment, focused on knowledge, would not be appropriate. A process-based assessment would take its place, and would include intellectual and practical skills as well as attitudes. Two problems were found: 'doubts about the degree to which scientific skills are actually transferable to all curriculum subjects', and the awareness 'that it might not be possible to ignore content completely' (DES/APU, 1989). In relation to the first, was agreed to distinguish three different *contexts* ('science', 'everyday' and 'other school subject'), a distinction however which was not much used as a variable to report performance. In relation to the second, it was decided to restrict any concept-dependence within the test questions to a short list of concepts of science.

The assessment framework developed by the Working Group, suffered a transformation. The new framework (called Science Activity Categories by the Monitoring Team) had an operational basis intended to meet the needs of defining test instruments and to report results in a form useful for teaching. This set of Science Activity Categories 'was not founded on any philosophical model of the nature of science. Nor was it based on any psychological model of stages of learning'. The categories were said, taken together, '... represent some possible and important outcomes of education, particularly in science, reflecting 'a view of science as practically-based investigative activity'. (DES/APU, 1989). Six Science Activity Categories were defined and appeared in time for publication in the first survey reports (pre-1980): 'using symbolic representations', 'using apparatus and measuring instruments', 'using observations', 'interpretation and application' , 'design of investigations' and 'performing investigations'. It is made clear that such categories are meant to represent identifiable different activities (involving a deployment of a number of intellectual and practical skills and abilities), but it was never claimed that they are mutually exclusive; 'indeed, there is some overlap between them in terms of the skills and abilities they demand' (DES/APU, 1989). It is worth noting the remark made by the Team in relation to one of the categories: '"Performing Investigations" enjoyed a unique status in that it was perceived and justified as the embodiment of an important aim in science teaching which encompasses more than the separate elements represented in the other Categories, all of which are involved in it' (DES/APU, 1989); in Woolnough's (1989) terms 'the whole is greater than the sum of its parts and different'. On economical grounds (the number of features manageable in a single survey), 'it was determined that the questions developed in any Category should be designed to be free of any dependence on taught science concepts..., except in the case of "Application" ' (using a list of concepts drawn up by the Unit)(DES/APU, 1989). Importantly, instead of imposing paper and pencil tests for every category, they 'agreed that where practical skills and abilities were concerned then these would be assessed in practical mode, using appropriate equipment and other physical resources as necessary' (DES/APU, 1989). An attempt to validate the framework of Categories (and sub-categories) was made; experienced science educators were asked to review it in some depth, in connection with its validity as a representation of those processes in science relevant in school education. They argue that 'consensus expert agreement is necessary to validate an assessment framework' and that 'empirical data cannot substitute for educational judgement' (DES/APU, 1989). Nevertheless, 'an appeal might in addition be made to empirical correlational evidence in support of the judgements made' (DES/APU, 1989). In relation to this stand, on empirical correlational evidence, the relevant observation is that: 'a poor correlation between performances on two groups of questions might support an assumption that the two groups measure different things, or might throw some doubt on a previous assumption about equivalence. A strong association between different test questions is not sufficient evidence that these do measure the same thing(s)' (DES/APU, 1989).

This remark seems to suggest that one needs to analyze the content of what is actually involved in each question (rather than the numeric value of a correlation), and I would add: the analysis of what is actually elicited from those questions.

In relation to the assessment of 11 year olds, attention will be focused on one Category: 'performance on investigations' (the six Science Activity Categories are broadly the same for all three ages). The relevance of this Category and therefore of the kind of tasks involved in it, is due to 'the requirement of the integration of a variety of scientific activities and the goal-directed investigatory or problem solving nature of the tasks presented to children result in *tasks which derive validity from being recognisably close to what might be happening in primary classrooms'* (my italics)(DES/APU, 1988). Assessment of 'performing investigations' was practical. Testers recorded on check-lists the way children carried out an investigation and asked them structured questions (to probe features of their performance which were not entirely observable: e.g. why certain actions had been carried out).

The team worked with fourteen investigations, finding evidence that process demands varied with content. Nevertheless, it seems that they found a way to make investigations to certain extent comparable when the say that 'it has also proved possible to make valid comparisons across investigations by re-analysis of check-lists' (DES/APU, 1988). A combination of statistical association, logical connection and educational relevance led them to postulate five components of the process of investigation ('each not necessarily present in every investigation'): 'general approach', 'control of variables', 'measurement of the independent variable', 'recording of findings' and 'nature of result' (DES/APU, 1988). It seems that there was no systematic attempt to represent science processes across different contents to see whether they are content dependent or not. Nevertheless, they are aware that task demand 'is inevitably expressed within a form, a subject matter, which to some extent interacts with and defines the particular nature of the process'; adding that 'although they have not been systematically investigated, effects of particular content have been noted in several areas of the framework' (DES/APU, 1988).

Techniques for the Assessment of Practical Skills in Science (TAPS):

This project has the aim of developing assessment instruments, to be used by secondary teachers, in evaluating or assessing practical work. In their attempt to revalorize practical skills and laboratory work, the project 'focused upon the non-trivial aspects of practical skills as they manifest themselves in the classroom' (Bryce & Robertson, 1985). Six practical skill categories according to coherent groupings of specific skill objectives coming from teaching materials were defined: 'observational skills', 'recording skills', measuring skills', 'manipulative skills', 'procedural skills', and 'following instructions'. Two additional skills were defined: 'inference skills' and 'selection of procedures'. The project seems to be influenced by a Bloomian view when it relies on 'specifying objectives for a particular course and the skills and abilities pertinent to practical work' (Bryce & Robertson, 1985).

In this attempt to revalorize practical science, (what is meant by this is not entirely clear) two problems seem to be mixed up: the nature of what would count as practical (in particular if cognitive abilities fall in this category), and a dissatisfaction that 'practical science' is usually measured by paper and pencil tests. There seems to be a hint of their position when they say that 'we do not intend to object the proposition that "practical investigations" are desirable educational targets. ...The latter [pupil's practical investigations] is much more complex than is fashionably supposed and there is little point in extrapolating wildly from the success of carefully structured tasks for the assessment of practical skills to what might be the most valid forms of assessment for practical investigations' (Bryce & Robertson, 1985).

Dutch National Institute for Educational Measurement (CITO):

Reports of the assessment interests developed in this Institute span across three articles: Hellingman, 1982; Alberts et al., 1986a; Alberts et al., 1986b. The Institute initiated a practical testing project in 1976 in order to help teaching staff develop the content and goals of an obligatory practical internal examination (in a perspective of having practical work as part of the final examination system) for upper secondary schools and the pre-university educational level. Hellingman (1982) reports the increasing demands for more teaching of experimental methods and expects the trend to continue. The practice in constructing objective tests used by the Institute has been, says Hellingman (1982), 'fairly closed in character' (using paper and pencil tests). Based on Kruglak's findings (1954), that paper and pencil tests 'are at best only crude approximations to the evaluation of ability to deal with laboratory materials and apparatus' Hellingman (1982) expresses the view that 'such tests would thus not seem to be acceptable measures for establishing pupil's competences in the practical, i.e. experimental aspects of science'. The shift to more open-ended educational situations is to be welcomed, but two main problems are pointed out: the lack of a clear framework to define what would be included under 'practical work', and standardization of the results (with multiple choice tests being not suitable for practical work, more expensive forms of setting standards have to be found). The rest of Hellingman's paper focuses on the criteria used to formulate a list of objectives for 'practical work': completeness and levels of generality. It is however not explained how the problems of deciding about completeness (what skills should be included), the nature of the skills or abilities, and the blurring of the distinction between practical and theoretical work, are resolved. The set of objectives is established in a sequence of what almost seems to be a handbook description of a procedure, from 'formulation of the research question' to 'report, in writing or speech, as it concerns' (Hellingman, 1982). The other papers describe different possibilities for practical assessment can be carried out (Alberts et al. 1986a), and the test construction procedure which was followed at the Institute (Alberts et al., 1986b), but no examples of the tasks to assess practical work are given.

From this closer review of some of the most relevant curriculum projects and assessment schemes for elementary school, the following general points emerge:

i) 'science processes' have been arrived at from a number of sources, including philosophical analysis, psychological or logical categories, or pragmatic and pedagogical consideration;

ii) open-ended practical tasks are becoming more widely used as pedagogical devices and as representatives of more valid assessment,

iii) there seems to be a lack of studies systematically exploring the dependence of science processes across different contexts;

iv) by adopting a Bloomian perspective of defining processes, proceeding a priori and by logical deduction, the opportunity to elicit children's own reasoning is missed; imposing a logical structure rather than a psychological and pedagogical one.

iv) in a Bloomian view of decomposing complex abilities (performing investigations for example) into all their allegedly required sub-abilities, is not very clear how performance in such abilities can be improved if they are learned separately;

v) developing curriculum material could possibly start by devising feasible, classroom-like, tasks whose cognitive demands are known, rather than subverting the logic of learning in favour of a logic of organization and assessment,

vi) the variety of understandings of 'science processes', 'scientific processes', 'process-skills', 'practical science', etc., may be partly explained by the lack of clarity in relation to the kind of constructs which are required to understand children's behaviour in science activities (e.g. performing investigations).

2.3 MAIN ISSUES IN RELATION TO 'SCIENTIFIC' PROCESSES IN ELEMENTARY SCHOOLS.

A critical model presented in Chapter 1 describes the main problems in relation to 'science processes', grouping them around problems of: a) transfer from the nature of science, b) defining and matching, and c) eliciting and discriminating. I shall now re-discuss some of the examples from the literature described above, raising issues about them from these three points of view.

2.3.1 PROBLEMS OF TRANSFER FROM THE NATURE OF SCIENCE.

There is little doubt, from the review, that curriculum projects and assessment schemes give a primary role to practical tasks as learning activities. Frequently, the core of such activities is what might be called 'science processes', although this category does not include a homogeneous set of processes. Science processes show a variety of conceptualizations, drawing on the nature of science and/or some psychological characteristics. One group of curriculum projects (SAPA, WSP) appears to be based on what may be called an inductivist-empiricist point of view in philosophy of science. Another group (Science 5-13, SCIS, SPACE) might be characterized by the emphasis given to more deductive procedures like making hypotheses and testing them, although sometimes mixing them with inductive ones. Finally others (APU, TAPS, CITO) may have adopted a more practical approach, influenced perhaps by summative assessment demands, or another (CSMS/CAT) that might even have a more 'rationalistic-psychological' (Piagetian) approach to processes.

Accepting that primary school science can be characterized to certain extent by its methods ('science processes') and setting aside the problem of their relation with conceptual change, there are several ways that 'processes' come into existence or are transferred to an educational context. The extreme cases (SAPA, WSP) are those which having identified some science processes within a particular context of philosophy of science, just transfer them to an educational context and in doing so take for granted that nothing essential has changed. These processes are then taken as the fixed framework within which to validate teaching and the selection of the curriculum. For example the case of SAPA, to validate the learning hierarchies (chain of 'sub-processes') within each 'process'. Similarly, and following as does SAPA the Bloomian approach, another project (Science 5-13), might select a less specific view of science, identify certain 'processes' or abilities, derive logical consequences for acquiring such complex abilities, evaluate them (represented as behavioural objectives) by their discriminating power, and conclude whether they 'exist' or not; finding out if there is learning by using similar activities or tasks. In a case where 'processes' are derived from a Piagetian or other psychological framework (CSMS), the processes are equally taken as given (their validity deriving from their source) and are not put into question by the results of children working on tasks derived from them.

There seem to be several problems that concern this 'transfer' of 'processes' from philosophy or psychology to educational practice; that is to children's behaviour on practical 'scientific' activities carried out in the classroom. One is the variety of inspirational sources that can be used: empirical-inductivistic, rationalist, or social-constructivist among others. And there are no a priori reasons why one should not use several of them at the same time (e.g. 'identifying variables', 'reasons for events', 'critical discussion of results'). Another problem is the context in which 'processes' are understood to be functioning. At least we need to make distinctions between practices in the context of discovery and in the context of justification within science, and also between the scientific and the pedagogic contexts. That is, the meaning or nature 'processes' may have in the context from which they are drawn, does not necessarily correspond to the meaning or nature they acquire within a pedagogical context. The lack of clarity in relation to these three different levels of analysis or practice, may explain the consistent demand for a 'clear framework' in practical science (Doran, 1978b; Hellingman, 1982; Bryce & Robertson, 1985) and the reported variety of ways in which 'science processes' are defined (Donnelly & Gott, 1985; Millar & Driver, 1987; Johnson, 1987; Jenkins, 1989; Fairbrother, 1989). Even cognitive psychologists have problems in defining cognitive skills and in offering a 'comprehensive and universally accepted theory' (Chipman & Segal, 1985a, 1985b)

Locating the level of analysis and practice for science processes at a pedagogical level would be one of the first tasks in science teaching for elementary schools. Locating 'science processes' at this level would imply, based on the distinctions made, differentiating between the practice of professional scientists and practice in the classroom, and also between philosophical categories in the scientific context of justification and the categories ('science processes') used to describe children's 'scientific' performance; helping to clarify the metaphor of 'children as scientists' behind the process approach in elementary science. Thus, science education would look for appropriate categories to describe and understand children's engagement in 'scientific' activities. In the case that some categories, derived from a philosophical context, are used in an educational context they would not be just transferred but suffer a 'didactic transposition' ('transposition didactique': Chevellard, 1985; Martinand, 1983); meaning that they would not necessarily have the same function and nature as in the original context. This argument leads us to consider defining processes as derived from practical tasks (e.g. doing investigations) which have been selected as being desirable to encourage; and not completely decided a priori, leaving room for determining processes out of what children can actually do on these tasks. This leaves open the opportunity to know about children's ways of reasoning, as has also been possible in the case of children's 'alternative conceptions', 'alternative frameworks', etc..

I have argued that one could define 'science processes' at a pedagogical level and that that would imply them being derived from desirable practical tasks to be encouraged in classrooms. This last concern is, in my opinion, one of the cornerstones for understanding the nature and role of 'science processes'; that is to consider the nature of the instrument to be used as means for eliciting processes. The nature of such an instrument is relevant in two ways: as a valid representation of what science is about at this level of schooling, and as an appropriate instrument to elicit children's 'scientific' behaviour. In the case of the Bloomian approach for eliciting 'processes', as in the case of SAPA and Science 5-13, 'processes' are defined a priori, just looking for appropriate psychometric properties before considering them to be functional. In this sense the pedagogical task becomes no more than an 'excuse' for eliciting what is already decided beforehand; the tasks are not primarily important as pedagogical activities by themselves, but as atomized instances of what it is intended to promote (usually logically derived 'sub-processes'). This explains, partly, why in SAPA etc. there are many tasks, one for each objective belonging to an intended 'terminal behaviour' (in this case 'process'). This derives from the intention of the originators to teach 'science processes' (SAPA, WSP), the method or methods of science, or 'formal operational thinking abilities'. With the Bloomian approach applied to 'science processes' a gap appears between 'processes', tasks and assessment; with 'processes' defined a priori as behavioural objectives, assessment as the evaluator of the intended curriculum material, and tasks as the instances that discriminate children's behaviour in terms of reaching or not reaching the *a priori* defined objectives. In this approach tasks are not relevant for themselves, but are chosen as a result of possessing psychometric properties and for their convenience for assessment purposes. This approach reflects more a rationale for organization and evaluation, than a pedagogical rationale based on what can be implemented and elicited with 'scientific' practical activities able to be carried out in the classroom, valued by teachers, and possible and interesting for children.

2.3.2 PROBLEMS OF DEFINING AND MATCHING.

Identifying and selecting 'science processes' from their original context, recognizing the differences and similarities between the context of origin and the pedagogical context, and deciding about the kind of practical classroom tasks that would validly represent scientific activity and elicit such 'processes', still leaves many problems unresolved. One is to define or represent the intended processes in tasks (practical or paper and pencil) or eliciting them from classroom activities, and another (not by any means trivial) is the question of whether processes are in any sense independent of content (for which matching intended processes across tasks is essential); such independence of content is seen as important by test developers (Johnson, 1987)

In the case of representing the selected processes in tasks, either using paper and pencil or practical tasks, there is the work done by curriculum projects and assessment schemes. Curriculum development projects validate their objectives ('processes' in this case) by constructing similar tasks to those designed to teach a science process, and test whether teaching a specific process to an experimental group makes a significant difference in comparison to a control group. If the difference between experimental and control groups favours the former, it is claimed that the ability or process has been transferred to a different content. The problem is that representing is sometimes restricted to a limited range of new content. As Shaw (1983) says: 'little effort has been expended on evaluating these programs in relation to the students' ability to apply problem-solving skills to situations or content not specifically covered in the curriculum being evaluated'. If the range of new content is limited, then the problem of representing the same objective or process across different contents is barely faced, being sometimes restricted to the phrasing of the items, without systematically reporting the problems of defining a process for different content. In the case of SAPA, which uses both practical and paper and pencil formats, no attempt to systematically define processes in different contents is reported (AAAS, 1968). And CSMS when testing the unity of formal operational thinking (consistency between schemas), Shayer and Adey (1981) report only three schemas being present in two tasks out of five, with five just in one task. The problem of defining processes for different content is at the heart of addressing the problem of whether independence of content can be claimed or not. Associated with the problem of representing processes in a limited range of contexts, is the problem of claiming independence of content for such processes; without empirical evidence is difficult to maintain such a claim. Nevertheless the impact of such claims even reached the APU work when, as described, there were proposals for processes across curriculum; although the Monitoring Team in science seems to have considerably reduced the importance of such proposals.

When processes are elicited from classroom-like activities, it is reported (DES/APU, 1988) that 'It has become evident that, depending on the particular content employed to frame an investigation, the burden of particular process demands has varied. These differences in emphasis were deliberately exploited rather than suppressed, question differences being maximised so as to provide evidence across the range of science processes'. This might be interpreted as at least some recognition of the difficulty, as also in 'Where the problem "defines" the independent variable, performance levels are high. Naturally enough, where some redefinition is required, levels are lower, as when two variables have to be manipulated in order to tackle the problem' (DES/APU, 1988). Obviously, the problems for different contents were defined differently in relation to 'general approach' ('setting up a situation in which the independent variable can be investigated') in performing practical investigations. From this piece of work it seems clear that more attempts should be made to systematically try to elicit the same processes across different contents and find out where the elicitation problems reside (Differences in cognitive demands perhaps?) and their relationship with certain kinds of content.

2.3.3 PROBLEMS IN ELICITING AND DISCRIMINATING.

Problems of eliciting and discriminating should not be seen as detached from the ones described in the prior two sections. The case of SAPA can be seen as paradigmatic. It not only has problems at the level of transferring 'scientific processes' from philosophy of science to education or defining processes in a limited range and claiming independence of content, both already explained, but also presents problems at the level of eliciting and discriminating.

Problems of elicitation appear when the intended process (with its performance criteria) is compared against children's actual performance on the process. In most curriculum projects (SAPA, Science 5-13) and assessment schemes, although not in all (APU), performance criteria are given to decide whether the objective or science skill is present or not. Thus, the job in curriculum evaluation and assessment consists in deciding whether (the usually a priori) behavioural definition of a process matches the answer given by the child or not. This, almost by definition (when multiple-choice tests are used), prevents us from knowing what children actually think. Certainly it helps to, as some researchers like Molitor and George (1976) do, interview children after the test (multiple-choice) to detect whether they are consistent or not in their answers; but this does not necessarily focus the attention of the researcher on the kind of answers children give, nor are they taken as important on their own. They report 'that not all items on each subtest were capable of eliciting the skill behaviour defined by the stipulated performance criteria. Certain inference items could be answered correctly without the stipulated skill behaviour' (Molitor & George, 1976). But where does the problem lie, in the science educators that judged the items as adequate elicitors of such behaviour, or perhaps in the type of answers given by children. Molitor and George (1976) suggest that 'Apparently, when the item has conceptual content of a high degree of familiarity a student will rely on past experience rather than on collecting data'. Nevertheless, it seems that more information about children's responses is needed to decide whether the definition of the process needs to be changed, or whether children's responses require a different description or judgement. Certainly with the help of performance criteria, the cutting edges are clearer and locate weaknesses in children's performance, but there is still the problem of how to understand, describe and, in the end influence that performance.

This, leads us to the problem of deciding what kind of instruments would be required to elicit what is decided on as to be encouraged in the classroom. In this sense arguments based on economical and practical grounds have prevailed, inclining people (McLeod et al., 1975; Molitor & George, 1976; Shaw, 1983; Ross & Maines, 1983; Tobin 1984) to use mainly paper and paper instruments in multiple choice mode to elicit science processes. The case of SAPA is unusual because, despite requiring written answers for filling in or selecting answers, children are required to perform practical activities; in contrast with the above mentioned trend. From Kuhn & Brannock's (1977) article it can be inferred that individuals tend to do relatively better using practical tasks than using just pictures in trying to isolate variables, suggesting that the format of the task affects performance. Only recently, a concern for using more valid activities (that is practical and open-ended tasks), seems to be growing (Hellingman, 1982; Alberts et al., 1986a, 1986b; DES/APU, 1981, 1988).

Where processes are elicited from tasks, as in the APU's 'performing practical investigations', there seems to be much more work to do. This may be a beginning of using more valid and classroom-like activities as a pivotal element between valuable behaviours to be promoted and information about children's improvement on performance. With instruments (based on open questions, open-ended activities, etc.) that allow for eliciting children's reasoning, it seems possible to hope to understand children's behaviour in their own terms and to be in a better position to understand how children change from stage of knowledge to more advanced ones.

In relation to discriminating children's 'scientific' performance, it seems first to be necessary to map children's behaviour in order to be in an better position to set up improved criteria to discriminate behaviour other than that which fits the description of the the intended process or not. Another problem that seems to need systematic study is the longitudinal investigation of processes (as cross-sectional studies were proposed with same processes). It might be the case that processes show maximum power of discrimination at certain stages of children's development, and less when they are either easier or more difficult, affecting the strength of their relationships.

The main conclusions of this review may be summarized as:

i) the problems described and discussed in relation to science processes, stress the view that they should not be taken for granted and that they can be analyzed at different and connected levels;

ii) there seems to be more demand and necessity for classroom-like practical tasks that encourage certain lines of development like "How do we know?", "How can we find out?" that have certain cognitive demands and set criteria for better judging children's abilities to undertake investigations.

2.4 PROBLEMS OF DEFINING AND ELICITING 'SCIENTIFIC' PROCESSES USING A MORE DETAILED EXAMPLE.

Although the main general problems about defining and eliciting processes have already been addressed, a particular example of how processes are constructed out of task analysis and children's performance would give a clearer idea of the problems faced in this field.

For this purpose, the work of the APU has been selected, given their leading role in trying to assess children's performance in science, strongly based on a process approach. Such an approach is seen by Paul Black (1990) as a major achievement of the APU. Emphasis will be focused on category 6 of their assessment framework, that is 'performing practical investigations' (DES/APU, 1988); which is a category regarded as 'essential' (DES/APU, 1988) and applied to 11 years old, the same population that is addressed in the present study.

2.4.1 DESCRIPTION OF APU WORK ON 'PERFORMING INVESTIGATIONS'.

Under this heading the APU team was trying to determine and assess the 'investigative skills' (DES/APU, 1988) children can put into use in doing practical investigations, as opposed to just giving written answers. Other process categories were also assessed ('use of graphical and symbolic representation', 'use of apparatus and measuring instruments', 'observation', 'interpretation and application' and 'planning of investigations'), but as the APU says, although these processes 'may be involved in carrying out an investigation, but the sum of the parts does not necessarily make the whole' (DES/APU, 1988).

Thus, the APU is interested in those 'spontaneous' behaviours (investigative processes) children show while performing practical investigations. To assess them, a number of practical investigations were designed. For this purpose, investigations were conceived holistically, this is, not predetermining the investigative skills required to perform the tasks; these were identified post hoc. This essential feature will be discussed in detail later. Children were given a problem to solve, and the job the APU set itself was to determine and assess the investigative processes involved in performing such problems. It is important to note that the problems or investigations developed by the APU can be considered as plausible classroom activities, as P. Black

(1990) says 'it was important to base the approach [APU's] to the work on classroom experience' and even further 'the new monitoring had to express the current professional views about the *best practice* ...' (my italics); although they were primarily designed for testing purposes. The complexity of such an attempt is recognized when they stress the problems faced in terms of logistics and validity (DES/APU, 1983, 1988; Black P., 1990).

A total of sixteen different investigations were designed between 1980-84. Attention will be focused on those used in the 1981 survey (six), a representative part of the whole set; although reference will be made to the 1982 survey. The 1981 survey provided an important opportunity to compare processes across tasks, because each of two groups of children performed one of two sets of three investigations.

Structure of investigations:

As stated above, these tasks were conceived as completely integrated problems for children to solve, leaving it to the researchers to determine the processes involved. The tasks are open-ended but have some degree of structuring. Each task has a clear objective: to find out about the behaviour of some phenomenon by practical manipulation of relevant factors (e.g. 'Find out which of these places woodlice prefer to be in'). It is up to children to decide how to tackle them. For the design of the tasks the APU relies on the idea of a fair test and on the control of variables (e.g. 'Test the papers in three different ways to find out which one would be the best to choose for covering a book'). The tasks are however partly structured by the introduction of questions and statements to be completed, intended to help children tackle the problem and to make their findings explicit (e.g. 'Put down here why the bulb did not light in the beginning'; 'Put down here what you find: The snails were fastest on ..., The snails were next fastest on ...'). Thus, tasks are a compromise between open-ended and structured investigations.

Task analysis:

Having designed the tasks, the next step was to make an *a priori* task analysis in order to provide testers with explicit check-lists for observing children's behaviour. Thus they arrive at a check-list containing a set of actions seen as needed to perform well in the task, from the point of view of the logic of the task analysis. Testers recorded children's behaviour by putting a tick or a cross for each observable check-point. In the case of the 'Woodlice' investigation for example, 27 check-points were used. Each analysis is content-related, that is, is specific for each investigation. Thus there still remains the problem of comparing processes across investigations. Indeed, as APU say 'There was no attempt to make the investigations similar to each other in their demand for investigative skills; they were designed to differ in their demands so that between them skills across the range required in scientific investigations were well represented' (DES/APU, 1988).

How these check-points were transformed into 'investigative processes' or 'components', comparable across investigations at least to some extent, will be discussed next.

Identification of 'components':

The identification of 'components' coming from the check-points identified for each task, derives from the worthy intention to 'summarise what the pupils did in ways which enable more generalised statements to be made about performance on various aspects of the process of investigation' (DES/APU, 1983), and therefore providing 'a way of comparing performance on different components within an investigation and on the same components across investigations' (DES/APU, 1983). The way in which 'components' were identified was by 'statistical procedures moderated by logical and educational considerations' (DES/APU, 1983, 1984). Check- points within each investigation were examined to see whether they related statistically to each other or not. Groups of check-points were identified where there was some statistical association between performance on check-points in the group. Final groups were then constructed, giving logical and educational considerations priority over statistical relationships alone. As a result, five 'components' or 'investigative skills' were constructed: 'General approach', 'Control of variables', 'Measurement of dependent variable', 'Nature of result' and 'Record of result'. These constitute what APU was able to achieve by way of constructing general 'scientific processes' intended to be valid for practical investigations. One may comment at once that of the five, only two, 'control of variables' and 'measurement of dependent variable' can be recognized as frequently occurring in various lists of 'scientific processes' (see earlier in this chapter). From its name alone, the meaning of 'General Approach' is hardly clear. The meaning and homogeneity of these components across tasks will be discussed in detail later.

Construction of scores:

Within each 'component' scores on check-points were combined (in various ways) to give an overall component score. The rules for combining scores made use of the relationships between the check-points within a group (DES/APU, 1983); in fact they claim to have found hierarchies of performance on certain check-points. For example, in the case of the 'Woodlice' investigation four check-points form the component called 'General approach': 'sets up 4 places, correct combinations', 'woodlice added after places set up', 'woodlice left undisturbed during trial' and 'woodlice closely observed'. The hierarchy found in this case is:

a) '4 places with correct combinations'
implies
b) 'woodlice put in after places set up'
implies
c) 'woodlice left undisturbed'
implies
d) 'woodlice closely observed'

Based on this, a score of 1 was given for the first check-point alone, and a score of 2 for the first plus two (or more) of the remaining three. The association between check-point a) being present or not, and the presence of all the remaining three or not, was statistically significant. This result, however, while consistent with the existence of the hierarchy claimed, does not in itself establish the detailed structure of the hierarchy.

The essential point, however, is that an attempt is being made to construct a component with a scoreable scale, by looking for hierarchical relations between check-points not initially defined for that purpose. If this could be clearly achieved for most components, the method would have evident advantages. In practice, this was not uniformly achieved, and scores of check-points were combined in a variety of ways using a variety of arguments and pragmatic considerations. Thus this leaves the coherent definition of the concept of 'components' in doubt, at the level of constructing scores for them. Nevertheless, the APU later on (DES/APU, 1988) recognizes this problem: 'the check-points used in defining a performance level for a particular component varied both in number and nature...'.

Comparisons within and across tasks:

When looking at variation of performance between components within each investigation, the APU finds 'different demands made by various parts of the process of investigation' and that 'the order of difficulty of different components within investigations showed few other patterns', concluding that 'the context (the subject matter) of the investigation had a strong influence on performance' (DES/APU, 1988). An alternative interpretation could be that the components differ from investigation to investigation.

Variations of performance on components across investigations were explained by APU in terms of the difficulties presented by the tasks. The question of whether components are similar enough across investigations to be compared in this way is stated but not (in the 1983 Report) resolved. 'Many further statistical analysis to be carried out on these data may throw light, for example, on whether high performance in a component for one investigation is associated with higher performance in the same component in the other investigations attempted by an individual pupil. The result of such analyses will be reported in a separate document' (DES/APU, 1983). I have not however been able to find any such analysis (e.g. in Report 3, survey 1982, DES/APU, 1984a).

2.4.2 CRITICAL DISCUSSION OF APU'S WORK ON 'PERFORMING INVESTIGATIONS'.

It is worth mentioning that their overall plan is consistent with the view put forward in Chapter 1, namely that it is a potentially useful idea to start developing classroom-like tasks and to try to elicit and construct 'science processes' (or 'investigative processes' `as the APU calls them) from them, and used as educational objectives, provided that they can be reasonably defined, elicited and measured. In this way educational objectives, pedagogical practice and assessment are brought nearer to each other; as Paul Black (1990) says, quoting the Task Group on Assessment and Testing, that 'assessment is at the heart of the process of promoting children's learning. It can provide a framework in which educational objectives may be set and pupil's progress charted and expressed'. Wellington (1989) says that 'inevitably in education the means of assessment determines the teaching style adopted'.

Design of plausible investigations:

In order to examine whether the APU investigations have some homogeneity of design, I have constructed a comparison between them, shown in Table 2.4.1.1.

This table show the objectives of each task as set, which I have then re-analyzed so as to see how far they can be expressed in comparable terms. It turns out that here there is some homogeneity: each task can be re-described as finding out how a dependent variable depends on certain factors to be manipulated.

However, the tasks appear to differ in the complexity of the combinations of values of independent variables to be manipulated. I shall call this the 'size of the search space'; an idea to be elaborated and used again later in the thesis. In the final row of Table 2.4.1.1 I have estimated the size of the search space, by multiplying together the numbers of values allowed for each independent variable - the number of combinations of values that there are. For example, in 'Woodlice', there are two values for each of two variables, giving a search space of size 4.

Tasks/ Features	Woodlice	Bouncing balls	Snails	Circuits	Paperback	Paper towels
- Objectives of tasks as set	'Find out which of these places woodlice prefer to be in'.	'Find out if the ball which bounces best on one surface bounces the best on all ...'	'Find out if the snails move just as fast on all the surfaces. If not, on which ones are they fastest? ...'	'Try to make the bulb light. When you have made it light find out what was stopping it lighting in the beginning'	'Test the papers in three different ways to find out which one would be the best to choose for covering a book'.	'Find out which kind of paper soaks up most water'.
- Find out (re-analyzed)	Find out how a dependent (dep.) variable (var.) (pref-erence) behaves in relation to 2 indep. var.	Find out how a dep. var. (bouncing) behaves in relation to 2 ind. var.	Find out how a dep. var. (speed) behaves in relation to 2 ind. var.	Find out how a dep. var. (lighting) behaves in relation to 3 ind. var.	Find out how a dep. var. (resistance of the paper) behaves in relation to 2 ind. var.	Find out how a dep. var. (paper soaked up) behaves in relation to 1 ind. var.
- Indep. variables (re-analyzed)	• lightness • dampness	• balls • surfaces	• snail's size • surfaces	• bulbs • batteries • wires	• different papers • different tests	• different papers

- Values of indep. var. (re-analyzed)	• light/not-light • damped/not-damped	• squash, golf, etc. • carpet, poly-styrene, etc.	• snail A, B, C, etc. • carpet, tile, etc.	• bulb changed/ not changed • battery changed/not changed •wires connected/ not connected	• wall paper, writing paper, etc. • no. of drops, no. of sanding strokes, no. of rubbing strokes.	• paper X, Y and Z
- searching space size (re-analyzed)	• 4	• 9	• 18	• 8	• 12	• 3

Table 2.4.1.1: Description and re-analysis of some features of the tasks of the 1982 survey.

If 'finding out' is re-interpreted in the way I describe it in the above table, then an important investigative process, 'searching', might have been (but was not) constructed, being the 'extent of the search' accomplished by children when performing such investigations. By comparing the number of combinations tried by children with the number allowed, one could have estimated how complete their investigations had been. Also, the size of the search space might be an indicator of the difficulty of a task.

It is also the case that some tasks (e.g. 'Bouncing balls' and 'Paperback') can be seen as requiring variables to be manipulated one at a time and thus emphasizing an investigative process like ' being systematic'. Indeed, this does seem to be reflected by some of the check-points: 'trials completed systematically', 'same kind of treatment all papers', belonging to the component 'General approach'. It should be noted that 'being systematic' in this sense is not the same as 'extent of search' as outlined previously. Both might have deserved study.

The demands of the tasks:

In order to fully understand the complexity of the problems faced by the APU team it is necessary to address the dilemma confronted by them. On the one hand, as mentioned before, they did not systematically try to elicit the same processes from each investigation; they accepted having ended up with a set of tasks having a range of cognitive demands, reflecting the particular content of each. On the other hand, they face the necessity to reduce the number and specificity of check-points, to fewer meaningful 'investigative processes', which could be looked for in different investigations, in order to make more general statements about children's performance. The problem is how to tackle both at the same time. The APU outcome will be clearer after the discussion of 'identification of components' that follows.

The task analysis generates a check list for each investigation. Unfortunately, a complete check-list is only given for the 'Woodlice' investigation; making it impossible to analyze them all and to follow the construction of 'investigative processes' from the check-points to their combination into 'components'.

Identification of components:

As stated above, the APU team used a combination of statistical, logical and educational criteria to transform check-points or task related actions into 'components'. Despite the fact that statistical criteria were not used in isolation, is unfortunate the correlations between check-points within each 'component' are not given in the report; thus making it impossible to judge the way the statistical criteria and others were used in combination.

The main point here, in relation to the construction of processes, is the elicitation of children's behaviour and its relation with the definition of a 'component' or 'investigative process'. In other words, how homogeneous is the elicitation process within each 'component'? How homogeneous in nature are different check-points within a 'component'?

In the following tables, the composition of the 'components' are shown. In the tables, I have also re-analyzed the 'components' in an attempt to see how much 'logical' coherence can be discerned in them.

Tasks/ Component	Woodlice	Bouncing balls	Snails	Circuits	Paperback	Paper towels
- General approach as set	• 4 places, correct combinations • woodlice put in after place set up • woodlice left undisturbed • woodlice closely observed	• surfaces used separately • trials completed systematically • same procedures used on all surfaces	• measures or compares speed acceptably • uses same procedure on all surfaces	• changes made systematically • battery X checked before or after discussion • effect of each change tested • battery X changed for Y	• same kind of treatment all papers (3rd test) • same kind of treatment all papers (2nd test) • same kind of treatment all papers (1st test)	• acceptable method of comparing or measuring soaked up water attempted • water added to paper (or vice versa) in acceptable way

- General approach (re-analyzed)	• setting up experim. conditions • fair-mindedness /organized • fair-mindedness /careful • observing	• setting up experim. conditions • being systematic • being systematic	• measuring • being systematic	• being systematic • checking • testing • identifying factors	• being systematic • being systematic • being systematic	• measuring • fair-mindedness /organized

Table 2.4.1.2: Description and re-analysis of the check-points for 'General approach' in all the tasks.

To illustrate the re-analysis, consider first the nature of 'General Approach' for 'Bouncing Balls' in Table 2.4.1.2. The second two check-points appear both to be able to be re-described more generally as 'being systematic'. Indeed, this seems to be the best re-translation of a considerable number of check-points in 'General approach' in the other investigations. However, we cannot say that 'General Approach' is simply 'being systematic'. The first check-point for 'Bouncing Balls' could possibly also be described in this way, but seems closer to something like 'setting up experimental conditions', which can be discerned also in 'Woodlice'.

In the end I found that I could reduce the number of types of feature in 'General Approach' to about seven:

- 'being systematic'
- 'measuring'
- 'fair minded'/'organized'
- 'observing'
- 'setting up experimental conditions'
- 'checking'
- 'testing'

One may not agree with such a list in every detail, but it seems difficult to avoid the conclusion that the check-points in 'General Approach', while having some coherence and homogeneity, cannot be covered by one simple descriptor. In particular, the interpretation is specially inhomogeneous for 'Woodlice'. Thus one has to conclude that on the face of it, 'General Approach' is clearer and more definite under re-analysis than it appears under the APU listing of check-points, but is nevertheless complex rather than simple in nature.

There are a few cases where two interpretations are given (see Table 2.4.1.2). Which is best depends on the kind of assumptions that one is prepared to maintain. If it is assumed that children know what they have to do in order to set the conditions of the experiment and start it, then 'woodlice put in after place set up' might appear as 'being organized'. But if it is assumed that a condition for a fair test is to expose the woodlice to all conditions without one of them taking advantage over the others, then the process might appear as being 'fair minded'; similarly with the other cases.

In the case of 'Control of variables', the analysis is as follows (see Table 2.4.1.3):

Tasks/ Component	Woodlice	Bouncing balls	Snails	Circuits	Paperback	Paper towels
- Control of variables as set	• places of equal area set up • intended equal areas • 6 woodlice or more used • woodlice placed randomly or equidistant	• intention to hold balls at same height • balls held at the same height • balls dropped, not thrown	------------	• checks that bulb A will work (after discussion) • checks that battery X will not work (after discussion)	• amount of treatment controlled for 3rd, 2nd and 1st tests • intends to control treatment for 3rd, 2nd and 1st tests	• whole area of paper well soaked • intention to take equal areas of paper • equal areas of each paper taken • initial quantity of water same • intention to soak all papers well
- Control of variables (re-analyzed)	• controlling observed (obs.) • controlling intention (int.) • controlling • controlling	• controlling (int.) • controlling (obs.) • controlling	------------	• controlling • controlling	• controlling treat. for 3rd, 2nd and 1st tests • controlling treatment for 3rd, 2nd and 1st tests	• controlling (obs.) • controlling (int.) • controlling (obs.) • controlling (obs.) • controlling (int.)

Table 2.4.1.3: Description and re-analysis of the check-points for 'Control of variables' in all the tasks.

'Control of variables' seems to be clearer in terms of definition than 'General Approach'; most of the check-points appear to elicit behaviours which intend to maintain certain conditions fixed while others change in order to evaluate the outcome for the dependent variable.

Despite this, 'control of variables' is absent from 'Snails' and the check-point headings ('checks ...') for 'Circuits' might be misleading. If 'checking' is interpreted as making sure of a finding, it implies more than just controlling, though this is also required. But 'makes bulb A light' or 'changes battery X for Y lighting bulb A' seem better descriptors and certainly require controlling 'wiring' and 'bulbs'. In the case of 'Snails', it is in the nature of the task that all snails are of different size (this variable is thus already given), so that it is impossible to control this variable in order to make fair comparisons of their speed.

One also has to point out that there is variation in the way 'controlling variables' is conceived: sometimes it is the children's observed behaviour, but sometimes the intention that they appear to have in mind (which might correspond or not to the previous one), and sometimes a treatment given in a test. This variation is however probably not serious. A reasonably good case can be made for this 'component' (in terms of 'logic') for 5 out of 7 investigations.

'Measurement of dependent variable' means basically how children operationalize (quantitatively or qualitatively) such a variable in the task. In Table 2.4.1.4 can be seen the variables that children are intended to operationalize, the check-points for this 'process', and my re-interpretation of them.

Tasks/ Component	Woodlice	Bouncing balls	Snails	Circuits	Paperback	Paper towels
- Operationalization of dependent variable* (re-analized) * not a component	• preference	• bouncing	• speed	• lighting** **already given	• resistance	• soaks up
- Measurement of dependent variable as set	• 4-7 minutes per trial • counts woodlice in each place • attempts to time interval of choice • trial timed or stopped when movement	• puts eye level with rebound • measurements accurate • measures height of rebound one at a time • attempts to measure using stick	• measures or compares actual path travelled • distance taken head to head or tail to tail • snail started from marked spot • end of movement marked • some measurement attempted	----------	CHOICE OF DEPENDENT VARIABLE IS THE EQUIVALENT COMPONENT HERE • testing for each treatment	• measures or compares water squeezed out • measures or compares water not absorbed • squeezes and/or drains papers equally •intention to squeeze and/or drain papers equally

- Measurement of dependent variable (re-analyzed)	• 'natural' measuring • indirect measuring • direct measuring • 'natural measuring	• internal measuring • accurate measuring • systematic measuring • external measuring	• measuring • fairness of measure • setting conditions for measuring • setting conditions for measuring • measuring	-----------	CHOICE OF DEPEN-DENT* VARIABLE • testing (choose) * It is in fact an independent var.	• measuring • measuring • fairness of measure • fairness of measure

Table 2.4.1.4: Description and re-analysis of the check-points for 'Measurement of dependent variable' in all the tasks.

'Measurement of dependent variable' (see Table 2.4.1.4) proved to be a difficult component to analyse. One example is the check-point: '4-7 minutes per trial' taken from 'Woodlice'. What does that mean? One interpretation might be that children are using a 'natural' way of timing or measuring woodlice preference for a place to stay in, by letting woodlice 'decide' whether or not to be in a particular place allowing them some time to do so; certainly more information is required from the APU's intentions with this check-point, as well as children's intentions with such behaviour, to have a clearer idea of what it would mean. Another example is 'puts eye level with rebound' from 'Bouncing Balls'. Here it seems that children make an internal and qualitative image of the height of the rebound of different balls, used instead of an external object to judge the phenomenon. One danger with keeping interpretations too close to a task analysis, as APU do in this case, is that such interpretations tend to reflect the analysis of the task (and the people behind it), but might lose certain features of children's performance.

Despite these problems, different aspects of 'measuring' seem to be present quite homogeneously in four tasks: 'Woodlice', 'Bouncing Balls', 'Snails' and 'Paper towels'. However, there is no dependent variable to be measured in 'Circuits' because the bulbs are either on or off, and in the case of 'Paperback' the dependent variable cannot be chosen, for the task only demands choosing an independent variable or test; while the problem with the dependent variable (resistance of the paper) is to judge its strength. Thus, a certain degree of homogeneity seems to be reached for four out of six investigations.

'Nature of result' and 'Record of result' are the other two investigative processes constructed by the APU.

Tasks/ Component	Woodlice	Bouncing balls	Snails	Circuits	Paperback	Paper towels
- Nature of result as set	• quantitative results recorded • results based on quantitative evidence • record at a) consistent with evidence • record at b) consistent with evidence	• quantitative results recorded at a) • results at b) consistent with evidence • results at c) consistent with b)	• quantitative details recorded at a) • result at b) justified in terms of measurement or comparison • result at b) consistent with evidence	• error identified correctly • bulb A made to light	• result at b) justified quantitatively • results consistent with evidence • results at b) consistent with other results	• quantity of absorbed water recorded • results justified in terms of measure-ments or compar-isons • result consistent with evidence
- Nature of result (re-analyzed)	+ Recording +Evaluating +Evaluating +Evaluating	+ Recording +Evaluating +Evaluating	+ Recording +Evaluating +Evaluating	• Identify factors • checking	+Evaluating +Evaluating +Evaluating	+ Recording +Evaluating +Evaluating

Table 2.4.1.5: Description and re-analysis of the check-points for 'Nature of result' in all the tasks (+ only accepted if a different interpretation of 'Nature of result' and of 'Record of result' is accepted as explained below).

One problem with these two components (see Tables 2.4.1.5 and 2.4.1.6) is that almost any recording made by children (demanded by the work sheet) is about 'results', making rather artificial the distinction between 'Record of results' and 'Nature of results'. From the educational point of view, one might prefer as an investigative process 'recording', whether of results or not, leaving open its content or nature. An additional process can be introduced to deal with 'results consistent with evidence' or as I suggest 'Evaluation', which in the APU scheme appears to be mixed with 'recording' in the component 'Nature of result' (see Table 2.4.1.5). In addition, 'Circuits' is at odds with the other check-points: 'error identified correctly' and 'bulb A made to light' do not look at all the same kind as the other check-points; the task ensures the results. According to the re-interpretation given, 'Nature of result' could be left with the job to elicit whether, what counts as a result, is a qualitative or a quantitative attempt. These distinctions might result in processes with a wider and clearer educational perspective.

The component 'Record of result' can be easily seen as coherent and homogeneous, as can be seen in Table 2.4.1.6.

Even so, this investigative process is absent from 'Woodlice' (despite a check-point: 'record made at a) and b) without prompting'), and 'test suitably titled' in 'Paperback' appears different to the other check-points; giving a suitable name to a test is substantially more than just recording.

The obvious reason why this 'process' is well defined is that it relates, not to an unobservable mental process, but to an observable physical process (writing down a result). Useful as including this outcome may be, it is not what most educators have in mind when they think of 'scientific processes'.

Tasks/ Component	Woodlice	Bouncing balls	Snails	Circuits	Paperback	Paper towels
- Record of result as set	------------	• record at c) without reminder • record at b) without reminder	• notes made at a) • results recorded at b) without reminder	• writes notes at a) without reminder • writes notes at b) without reminder	• results noted after each test • results noted at b) without reminder • test suitable titled	• record or notes made at a) • results written at b) without reminder
- Record of result (re-analyzed)	------------	• recording • recording	• recording • recording	• recording • recording	• recording • recording • 'naming tests'	• recording • recording

Table 2.4.1.6: Description and re-analysis of the check-points for 'Record of result' in all the tasks.

Overall, then, although APU could achieve some degree of consistency in the combining of check-points derived *post hoc* from classroom style tasks, they can not be said to have wholly succeeded. There is always at least one investigation for which each 'component' is problematic or missing, and in every case there are other anomalies. The component 'General Approach' seems best, but not perfectly, described as 'being systematic'. As for the structure of the tasks, an additional and pertinent process, 'extent of search', might have been introduced.

2.4.3 SUMMARY.

With the analysis of the APU work at a more detailed level it is hoped that the problems of defining and eliciting 'scientific processes' have become more transparent.

From this review it might be concluded that:

i) it is difficult to get consistent and homogeneous processes; they require an empirical process to make it possible to construct them, and they demand a careful blend of statistical, 'logical' and educational criteria in their construction.

ii) without a correlational study of children's performance across tasks, it is difficult to judge whether a consistent and homogeneous 'logical' definition of a process, actually can be considered as a dimension of children's performance to be measured; if one wants to get useful and more general descriptors and statements of children's performance than task related descriptions.

iii) the presence of processes need not be restricted to more or less traditional ones (those in almost every list of 'processes'), when there are signs that others like 'extent of search' look feasible to be constructed.

iv) more processes might be required to give a full account of children's performance on practical investigations, since only 5 processes were identified (one of which seems unclear in nature: 'General Approach') and with two of them ('Nature of Result' and 'Record of Result') seeming to mix children's spontaneous records of results with specific records requested of them in the task itself.

v) despite all this, the APU work suggests that by eliciting processes from practical tasks, one can hope to get the relationships between designing classroom-like activities, establishing educational objectives and assessing children's performance, in a closer relationship to one another.

Chapter 3.
RATIONALE AND STRUCTURE OF TASKS.

This chapter will introduce the tasks used to elicit children's 'scientific' behaviour and give evidence of teacher's opinions of the tasks, including their difficulty and the importance of 'processes' within them. But before embarking on this job, it seems necessary to outline the rationale upon which the structure and selection of the tasks and the design of the study are based, and to stress that the tasks developed can well be seen as curriculum development products as well as means to elicit processes; aimed at the goal of getting pedagogical practice, educational objectives and assessment nearer to one another.

3.1 RATIONALE FOR STRUCTURE AND SELECTION OF TASKS.

3.1.1 RATIONALE OF TASKS.

Having in the previous chapter reviewed several curriculum projects and assessment schemes concerning 'science processes', including the pioneering APU work on designing practical investigations (classroom-like activities with assessment purposes) based on a process approach framework (see Section 2.4), a beginning can be made on constructing a rationale for the present research, with the following remarks:

- there is a case for developing tasks to elicit a wider range of processes than were identified by APU, so as to widen and deepen our understanding of children's performance as well as our understanding of the problems of constructing such tasks;

- there is a case for trying to elicit the same processes across different tasks, so as to critically examine the problems of combining statistical, 'logical' and educational criteria in the construction of processes;

- there is a case for examining the consistency of performance across tasks by individuals, so as to examine the possibility of making statements about children's performance about relevant educational objectives, and about how far children's performance is content-related;

- there is a case for studying the elicitation of processes both in unstructured holistic investigations, in which children make all the essential choices, and in more structured tasks where the processes to be elicited are more explicitly planned.

The present study attempts to address this range of issues.

These considerations lead to a rationale for the research, as a study, not primarily of how well children perform on various processes in various conditions, but of the problems involved in constructing meaningful tasks to elicit processes at all, and of how far such processes can reliably be elicited in different ways in different contexts. In the process it is hoped to construct tasks which will prove viable and useful for classroom use in Primary Science. Thus the research focuses on the critical examination of the notion of "eliciting scientific processes through tasks of a kind and content suitable for children". General arguments for this position were given in Chapter 1.

Taking into account the remarks above, the main features of the study which seem to be needed are:

1) providing tasks of a holistic, unstructured nature,

2) providing other tasks structured so as to elicit processes previously determined,

3) a good range of processes to be elicited,

4) several different contents for each task,

5) unstructured and structured tasks to be paired to have the same content,

6) processes to be studied in each type of task to be the same for all the contents,

7) individual children to attempt tasks of each type with more than one content,

8) if possible to have a task which is more or less 'content free', so as to study how far this notion makes sense.

The tasks in 1) and 2) are in a sense complementary. By definition, in an unstructured task, where children are not guided to certain kinds of response, the nature of the processes which can be elicited (without interfering) are limited. Thus the structured tasks permit at least the possibility of understanding better how children perform on the unstructured tasks. They also allow the possibility of finding out something about the children's knowledge of the relevant domain.

The design and construction of tasks in fact proceeded iteratively. Possible content areas, common in Primary Science, were selected. A uniform simple structure for the unstructured tasks (to be described below) was decided upon, and tasks built around it. Video recordings were made of children attempting these unstructured tasks, in pilot work. Lists were drawn up of potential processes to be elicited in parallel structured tasks. Attempts were then made to construct such structured tasks in each content area, parallel to and using the same materials as in the unstructured tasks. This led to a refining of the processes and of the tasks to elicit them, so that there was at least some a priori reason to see them as similar in all tasks.

A decision had to be made about the nature and number of content areas to be studied. The minimum number, to study differences between content, is evidently two, but this makes the risk that the contents selected will be special in unknown ways, too serious. In the event, four content areas were chosen, one however being 'content free'. Given the decision to have two types of task for each, this brings the number of task-type combination to eight, which is already large enough to restrict seriously the number of children who could be studied. The number of children participating was 24, with all children attempting all tasks in both versions, making a total of 192 sessions. This overall design was chosen so as to maximise information about differences in content and in type of elicitation. The decision to elicit several different processes within each approach, using four different contents, reduces the number of children able to be studied performing the tasks to the point where the price of not being able to make any claims of inferences about the population has to be paid. But it does allow the exploration of the main problems faced in this field: constraints in defining and eliciting processes, comparability between processes across contents, and consistency of individuals performing in different contexts.

3.1.2 CRITERIA FOR THE DESIGN OF TASKS.

To elicit children's 'scientific' behaviour the tasks had to meet certain criteria:

- *Scientific content:* tasks should deal with some of the main phenomena that can be found in science, and which are relevant to this age group.

- *Diversity:* tasks should cover a reasonable range of different topics in science to get an adequate sample of children's 'scientific' behaviour, and to permit investigation of whether or not the tasks systematically elicit similar processes across different content. One task without specific 'scientific' content (Black Boxes) was included to see whether the behaviour in such a task would be similar to others that had a similar structure but 'scientific' content.

-Practical: the tasks were to be practical, as opposed to just paper and pencil activities, where situations or problems are only mentally represented. They would resemble, as much as possible, a 'real' situation where children manipulate apparatus, notice things and write notes or thoughts. They should as far as possible (particularly the unstructured) be plausible as ordinary classroom activities.

-Intellectual: they should require some level of reasoning, either to conduct the open-ended investigation or to answer structured questions related to some 'scientific' process. Manual dexterity and pure sensory discrimination should not be enough to deal fully with the tasks.

- *Simple:* the tasks should be easy to be set up inside the classroom and not require sophisticated equipment. At the same time they should not be trivial to do: each was designed to require children to manipulate at least two factors or variables.

- *Individual:* the tasks should be performed individually, although the potentiality of the tasks as group activities is recognized.

- *Suitable :* tasks should be able to be performed successfully by 4th years juniors in primary schools.

- *A uniform structure:* tasks of the same form (unstructured or structured) should have as far as possible a parallel structure across different content, in particular eliciting if possible similar kinds of process.

The unstructured tasks should place emphasis on children's spontaneous performance. Since continuity is the key element for preserving spontaneity, such tasks should allow for little or no interruptions at all. As explained above, the structured tasks were to be complementary, obtaining further information about children's thinking and behaviour. It was essential that they share the same content and that the same set of apparatus was used for both.

3.1.3 TASK CONTENT.

In the event, the content areas for which tasks were developed and used in the research were:

- rolling a ball on a sloping surface ('Rolling' = R)
- using a simple balance with a set of bricks ('Balancing' = B),
- floating a weighted straw in different solutions ('Floating' = F),
- identifying the contents of a 'black box' ('Black Boxes' = BB).

The first three feature frequently in work proposed for Primary Science. The last is a version of a rather well known 'content free' task. The tasks differ in the background knowledge they require. As a set, they seemed to satisfy the criteria in Section 3.1.2 reasonably well.

3.1.4 INVESTIGATION PLUS GOAL TASKS.

The form chosen for the unstructured tasks was 'Investigation plus Goal'. The broad concept of such a form was adapted from a proposal by Nedelsky (1965) for practical assessment tasks for university students. In Nedelsky's tasks, the student is told a goal: for example that he must be able to predict how far a ball rolling down a slope will run along a carpet. But the student is not told the conditions under which the goal is to be tested - for example where the ball will start on the slope. All the student knows is that the teacher will set some conditions, and that the student's quantitative prediction will be asked for and then tested.

In adapting this idea for primary children, the goals were simplified: to get a balance level; to get a straw to float at a given level in salt solutions; to fire a ball on a slope so that it entered a trap; to identify one closed box out of four. The child is required to investigate different ways of achieving these goals, so that when the experimenter chooses one set of conditions (weights and their position to balance, which solution to use, where to put gun and trap, which box to identify), the child can reach the goal.

This version of the tasks was intended to demand a range of 'processes' or behaviours: noticing relevant phenomena, identifying factors that affect other factors, replicating results, controlling variables and forecasting outcomes, all processes which bear some relation to what scientists do when engaged in doing an investigation. At the same time, it was hoped that this range of 'processes', behaviours and activities, would be displayed in an integrated effort driven by the aim of solving the problem.

As explained above, the tasks were designed in such a way that they all would have broadly comparable goals to achieve, so that one could hope to make valid inferences from comparing children's behaviour across them.

Thus it was the aim of these tasks to allow children to display 'processes' as best they could, and to be driven by the intrinsic motivation of the problem, instead of by the explicit desire of the researcher to know about how they would tackle it.

3.1.5 STRUCTURED TASKS.

These versions of the tasks were designed to give information about specific items of knowledge and processes, relevant to but not available as data from the Investigation plus Goal tasks. It was found possible to construct them all around the same six 'processes': noticing, understanding, 'what-if' reasoning or predicting, identifying variables, making generalizations and identifying causes. Responses in this form of the tasks are under explicit request, in clear opposition to the Investigation plus Goal forms of tasks, where actions are assumed to be spontaneous, guided only by the goal.

3.2 THE FORM GIVEN TO ALL TASKS.

The common structures, rationale and procedures for the two types of tasks will now be described, followed by the materials required for them and a description of each task; as well as the list of processes to be elicited and their broad definition in the tasks.

3.2.1 INVESTIGATION PLUS GOAL TASKS.

3.2.1.1 Structure of tasks.

All tasks in this version have three elements:

a) *Goal* : all tasks have a clear and specific goal. The tasks were constructed in such a way that the goal can be reached in different ways, being up to the children to decide how to do so.

b) *Investigation period* : children were allowed a period of about 15 minutes to achieve the goal of the task in as many ways as they could.

c) *Problem* : at the end of the investigation period a problem is set, of achieving the goal in conditions chosen by the researcher, and not previously known to the child.

- *'Processes':*

The investigation period is intended to elicit the natural 'processes' that children tend to use when left with the option of deciding by themselves how and to what extent to investigate a problem in looking for solutions to the task. The only restrictions are imposed by the characteristics of the tasks, and will be explained when the tasks are described in detail. Several 'processes' were looked for:

1) Extent of search.

2) Controlling variables.

3) Replicating or checking.

4) Making notes.

In fact, of course, being able to study such processes was contingent on finding ways of recording and noting children's behaviour, without interrupting their work.

Extent of search:

By keeping records of which variables were changed, and the values they were given, it was possible to see how extensively the children searched through all the possibilities offered by the situation.

Controlling variables:

The idea was to see how many changes of variables children introduced when doing the task, and whether they tend to introduce more or fewer changes after success or failure; giving an idea of how systematic they were.

Replicating or checking:

To see whether children made sure or not of their results, by repeating them. Differences might be expected after success or failure at reaching the goal.

Making notes:

The use of external memory is essential for many complex tasks. Thus making notes while performing the task, was thought of as a possible behaviour to be looked for (children were provided with a sheet of paper and pen and told they were there if needed).

3.2.1.2 Procedure used for tasks.

Children were individually given a sheet of instructions containing also a picture of the apparatus. They were asked to read the instructions and were then asked what the goal was and what they would do at the end of the investigating session, to see if the instructions were understood. It was emphasized that they should try to achieve the goal in as many ways as possible. If any problems about these or other matters arose, an explanation was given trying to keep as close as possible to the original instructions.

To minimize problems due to manual dexterity, children were allowed a few minutes to operate the apparatus and to feel some confidence in using it.

They were then told that they had 15 minutes to find out all the ways they can imagine to reach the goal. When 1 minute was left they were warned about the time and asked if they wanted to continue or not, and if they wanted to continue they were allowed to do so until they were ready. When they had finished, the problem was given, and their prediction was tested.

In order to record the children's behaviour, an observation check list was developed for each task (see below). These check lists try to capture the actions taken by children in terms of factors or variables altered, recording also success or failure in reaching the goal (see Instruments, Appendix A).

3.2.1.3 Description of tasks.

Balancing (see Illustration 1, Appendix I):

- *Materials:* + A modified Osmiroid™ balance.
  + A set of modified Lego™ bricks in different sizes and colours.
  + A sheet of paper and pen.

The balance was modified in order to allow the possibility of using different distances as well as different weights. The bricks (of the same size) were made equal in weight by adding some plasticine and covering them underneath with paper.

- *Description:* The balance had 4 stripes (A B C D) on which to put bricks. There were three different kind of bricks (small, medium and large) to be used. Thus a fair number of different ways to get the balance level can be found by combining such features.

- *Goal:* Get the balance level.

- *Problem:* Get the balance level with 2 medium and 2 small bricks on stripe B, having to use stripe D (children were asked to say which bricks would get the balance level first, before trying them).

Floating (see Illustration 3, Appendix I):

- *Materials:* + Straws.
  + Ball bearings.
  + Plasticine.
  + Salt.

+ Beakers.

+ Water.

+ A sheet of paper and pen.

- *Description:* Three equal floating straws were made by putting plasticine on one of the ends and marking them at every 0.5 cm, with a red line (Plimsoll line) as point of reference to judge their floating levels. Five different concentrations of salt solution were provided (400 ml of fresh water (A) and 12.5 g (B), 25 g (C), 50 g (D), 100 g (E) of salt in water). These were called 'different solutions' or 'different kinds of water'. The solutions were visibly different (cloudiness). If children asked they were told there were different amounts of salt in the water. By varying the number of ball bearings in the straws, they could adjust the floating level. Two straws (one always floating with its Plimsoll line above the water and the other below the water level) were already made for them to try.

- *Goal:* Find out how much steel shot to put in, to make a straw float with the line just at the water level in any given solution.

- *Problem:* How much steel shot is needed to get the Plimsoll line at water level in one of the beakers selected at random.

Rolling (see Illustration 4, Appendix I):

- *Materials:* + Inclined plane (board).

+ Marble.

+ 'Gun' (a modified dynamometer).

+ A wooden trap (U shaped).

+ A sheet of paper and pen.

- *Description:* The board was divided into two sections: a 'shooting' area (a band 15 cm wide on the left hand side) and the rest where a trap was allowed to be placed. It was set in such a way that the upper edge was 10 cm above the lower. Children should place the 'gun' anywhere in the 'shooting' area on the left and the trap wherever they wanted out of this area. The 'gun' was already graduated, if children wanted to know how much 'power' they were using.

- *Goal:* Get the marble inside of the trap.

- *Problem:* Get the marble inside of the trap with the 'gun' half way between bottom and top, and the trap in the right-bottom corner, choosing velocity and direction. Here the forecast was made by making a drawing before attempting to shot the marble into the trap.

Black Boxes (see Illustration 2, Appendix I):

- *Materials:* + Five wooden boxes.
+ Five marbles.

- *Description:* All the boxes had wooden walls inside forming different patterns, each with a sliding lid. A marble inside could roll but not jump over the walls. Four boxes were used in the problem, with the fifth used for them to see what a box might be like inside.

- *Goal:* Find out which box was which (by rolling the marble inside the closed boxes).

- *Problem:* find out what the pattern was in one of the boxes selected at random.

3.2.1.4 Observation records.

Children doing the Investigation plus Goal tasks were observed and their actions were recorded on detailed check lists (see Appendix A). Each check list had across the top all the variables the child could manipulate, and their possible values (made discrete if necessary). A tick was entered for each value of each variable chosen in any trial. There was also a column for success or failure of a trial. Time ran down the side of the sheet, with one time interval defined as a trial or attempt in which variables were modified (or not) leading to a success or failure at the goal of the task (e.g. balancing). Thus the record showed the complete time sequence of the child's actions with the apparatus, together with their effect at each attempt.

3.2.2 STRUCTURED TASKS.

3.2.2.1 Structure and description.

The Structured tasks used the same materials as in the Investigation plus Goal versions. They are described below in terms of the 'processes' they were desired to elicit, showing how these are represented in each task.

All tasks were divided into six sections, each dealing with one 'process'. These were (in the same order as presented to children):

1) Noticing.
2) Understanding.
3) 'What-if' reasoning or predicting.
4) Identifying variables.
5) Making generalizations.
6) Imagining causes.

Each 'process' is elicited mostly by two, but sometimes by three different questions or sub-tasks. In most questions children have first to do something and then answer a question. Some questions (mainly 'what-if' reasoning and 'identifying variables') asked for a justification, as well as prediction or identification.

An attempt was made to elicit evidence of the 'same' six 'processes' in all four tasks. Rather than starting with a tight definition of each process, and attempting to design the activity to fit it, it was decided to start with a less rigid definition and to try to adapt it to the constraints imposed by the tasks themselves; trying to compromise between something pedagogically meaningful and something well defined a priori.

Such compromises started from the following guide-lines:

Noticing:

This 'process' essentially concerned noticing or perceiving changes or differences. To generate something to notice, children were asked to produce phenomena such as: rolling the ball on the board from different positions, putting the straw with same weight in different densities, or adding weight within the same density, getting the balance level or tilted in certain conditions, and rolling the ball in different boxes (see Instruments, Appendix B).

Understanding:

This 'process' concerns how children understand certain phenomena or the effects of certain factors. In Rolling, Floating and Balancing, children were asked what kind of paths were possible on the inclined plane, what would affect the floating level of the straw and, what would make the balance become level (see Instruments, Appendix B). With Black Boxes eliciting this process was difficult, because this task is free of obvious 'scientific' content. I decided, then, to use this part of the task to get information connected to the kind of strategies children are willing to use in order to solve the problem posed by the Investigation plus Goal task (see Instruments, Appendix B).

'What-if' reasoning or predicting:

Here, it was intended to elicit children's ability to forecast an outcome, given certain conditions; in other words, what would happen if... certain conditions were met, that is the ability to establish functional or causal connections between variables. The neatest cases were Balancing and Floating. In the case of Rolling I decided to choose place for 'gun' or trap, and to ask children to select a place for the other and to predict the path of the ball towards the trap. In the case of Black Boxes children are asked to forecast a discrimination or its absence given certain (hypothetical) conditions; because there are no variables but tests it was called hypothetical reasoning. They were also asked to set, practically, the ball in all four corners - the first condition to solve the task - and to find a way to decide which box was which, that is to devise a test (see Instruments, Appendix B).

Identifying variables or factors:

For this 'process' children are asked what would affect an event, such as: identifying the factors affecting the ball going inside of the trap, the balance being level, or the floating level of a straw (see Instruments, Appendix B). For Black Boxes, the nearest thing to a variable was 'what to pay attention to' so as to distinguish between boxes (making a test) (see Instruments, Appendix B).

Making generalizations:

Children were reminded that some results can be obtained in different ways: e.g. getting the balance level or getting the straw at a special mark. The questions attempted to elicit what conclusions children could draw, by asking them what remains the same in these cases (see Instruments, Appendix B). For Rolling, children were reminded they have seen different paths for the ball and then asked what remains the same in all of them (see Instrument, Appendix B). In the case of Black Boxes, with pictures of all boxes in front of them, they were asked what would be the same in rolling the ball around all edges. Although these questions arguably represent generalizations, they differ somewhat from generalization in the sense used in the case of Balancing and Floating.

Imagining causes:

The intention, here, was to elicit ideas about why things happen as they do. Thus children were asked to explain certain events: why the ball goes as it does after two different rollings on the board, why two different agents (bricks and fingers) produce the same effect on the balance and why the ball can follow different paths inside of the boxes (see Instruments, Appendix B). The case of Floating was more complex. One question asks about why things float or sink. In addition children were asked what keeps the straw up, and whether a liquid that makes everything float in it could exist (see Instruments, Appendix B).

Chapter 5 describes how scores for such 'processes' were constructed, together with the inherent problems of doing so, and Chapter 7 looks at the degree of success of the attempt.

The fact of having similar 'processes' in all the tasks, allows looking at the relationships between processes within each task, and at their relationships across tasks, so studying the degree of content independence of processes. At the same time, as explained above, processes were not rigidly pre-defined, thus allowing a critical examination of the idea of designing different tasks to elicit the 'same' processes.

3.2.2.2 Procedure used for tasks.

Children were shown how to use the apparatus, as for the case of the Investigation plus Goal tasks. Then they were given the first page of the instrument (all tasks were 5 pages long) and asked to read and then answer the questions. I made clear that they could ask about any doubt they might have. The subsequent pages were given out one at a time.

3.3 WHAT DO TEACHERS THINK ABOUT THE TASKS?

It is known from previous work (Black, Harlen and Orgee, 1984) that teachers (and others also) can be poor judges of the difficulty of tasks, with variations between their judgements being much larger than children's actual variation in performance. For this reason, and so as to have it possible some reference against which to judge children's performance, it was decided to ask a group of teachers to predict performance on the Investigation plus Goal tasks. Also, given the necessary ad-hoc nature of the processes actually elicited in the research, they were asked to rate the processes as objectives.

The intention was to get some kind of indication as to whether the Investigation plus Goal tasks could be considered pedagogically suitable, and as to whether the processes present in the Structured tasks could be considered as relevant educational aims for children ending primary school. It was important to know how teachers expected children to perform on the tasks, whether they were prepared to use them at school and their reasons for doing so, in order to have an indication of the validity of the tasks. And in relation to science processes, I considered it relevant to know how important some of them were in terms of behavioural categories to be promoted and how difficult they were considered to be to cope with. Teachers were shown the actual materials for each Investigation plus Goal task and I demonstrated what was required from the children. They were also given a copy of the instructions given to children in the tasks.

The sample consisted of sixteen primary school teachers with a varied range of experience in primary school science, attending a course at the Institute of Education addressed to teachers in charge of science in primary schools. Thus the group was by no means a random sample of ordinary primary school teachers; on the other hand they would seem to be people who might have thought-out views about the suitability of the tasks, the importance of some processes and their level of difficulty.

- *The instrument:* (See Instrument, Appendix C).

The first two sections deal with teacher's expectations of children's performance in the Investigation plus Goal tasks, whether they would use them, and whether the teachers see the tasks as involving children in reasoning. The other two deal with teachers' expectations of performance on processes and with their rating of some of these processes as educational aims.

- *Results:*

Teachers expect children not to fail on the Investigation plus Goal tasks, although the results might be misleading; depending on how they are interpreted. If 'do well' and 'in between' are put together, these two responses account for around 80% of the expected performance in all tasks (Figure 3.1).

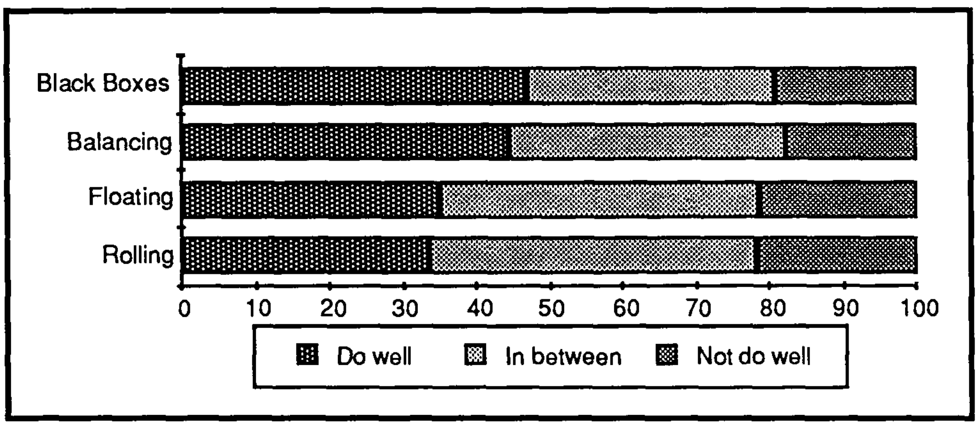


Figure 3.1: teacher's expectations of children's problem-solving performance by task in percentages.

This, suggest that teachers believe that a 'fixed' percentage of children would fail, regardless which task. Assuming that this belief is true, then the expected difficulty in performing the tasks would be best judged by the ratings given to 'do well' only. Doing so, Black Boxes and Balancing are considered easier than Floating and Rolling. When asked if they would use the Investigation plus Goal tasks as teaching devices, the great majority of them (14/16) would use either all or several (see Figure 3.2).

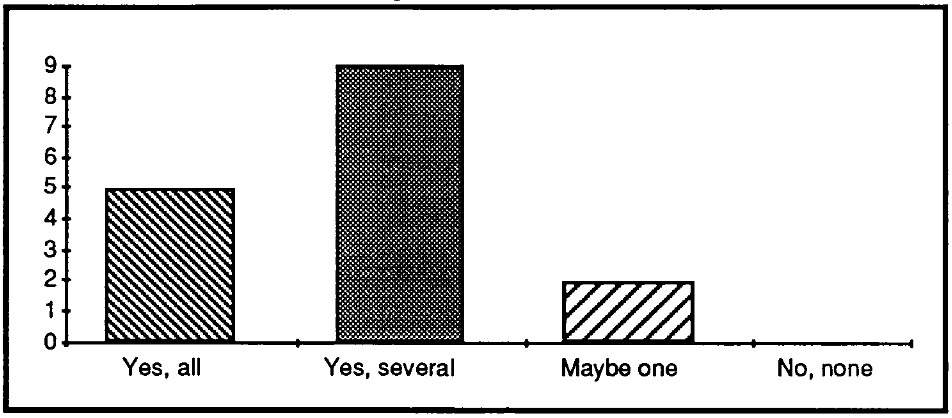


Figure 3.2: amount of teachers that would use the tasks as teaching devices.

And when asked about the reasons why they would use such tasks, just over half (9/16) gave reasons based on the encouragement children would receive for using logical thinking or scientific processes (see Figure 3.3), with almost all of them (14/16) considering the tasks as 'fairly difficult' (see Figure 3.4).

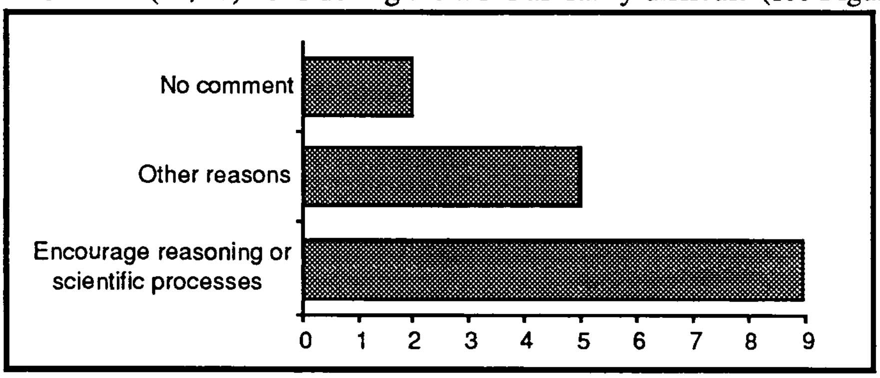


Figure 3.3: teacher's reasons why they would use the Investigation plus Goal tasks.

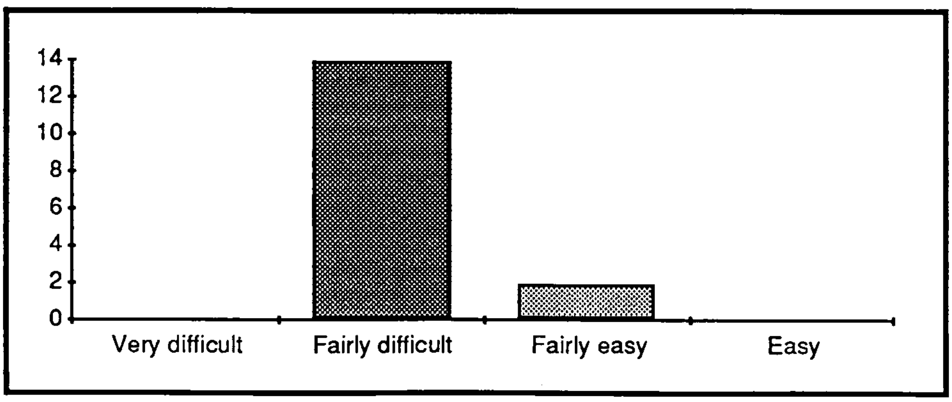


Figure 3.4: teacher's rating of Investigation plus Goal tasks difficulty.

In relation to scientific processes, if 'well' and 'in between' are taken together, the expectancy of coping with the situation is as follows: nearly all children (90%) are expected to do so for 'noticing changes', a great majority (80%) for 'setting conditions' (kind of predicting) and 'identifying variables' and , just above half of them (70%) for 'making generalizations' and less than 70 % for 'imagining causes' (see Figure 3.5).

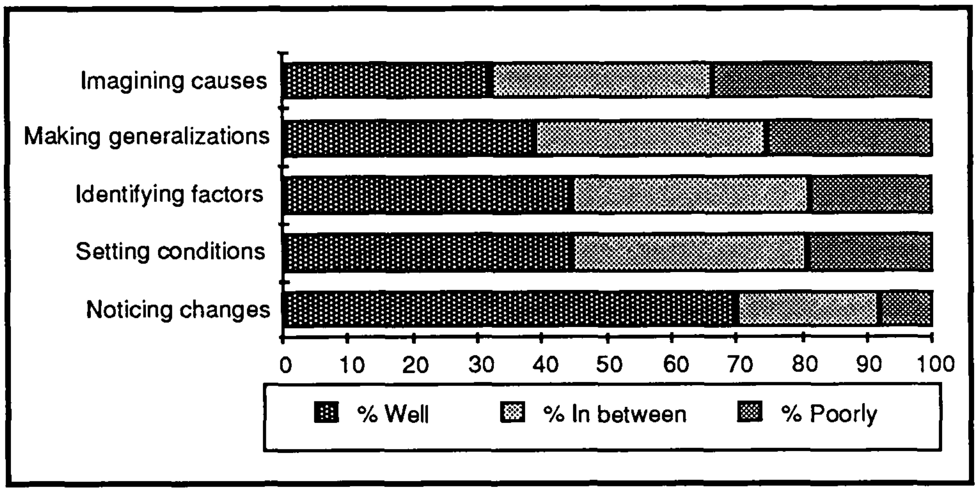


Figure 3.5: teacher's expectations of children's performance on certain scientific processes by percentages.

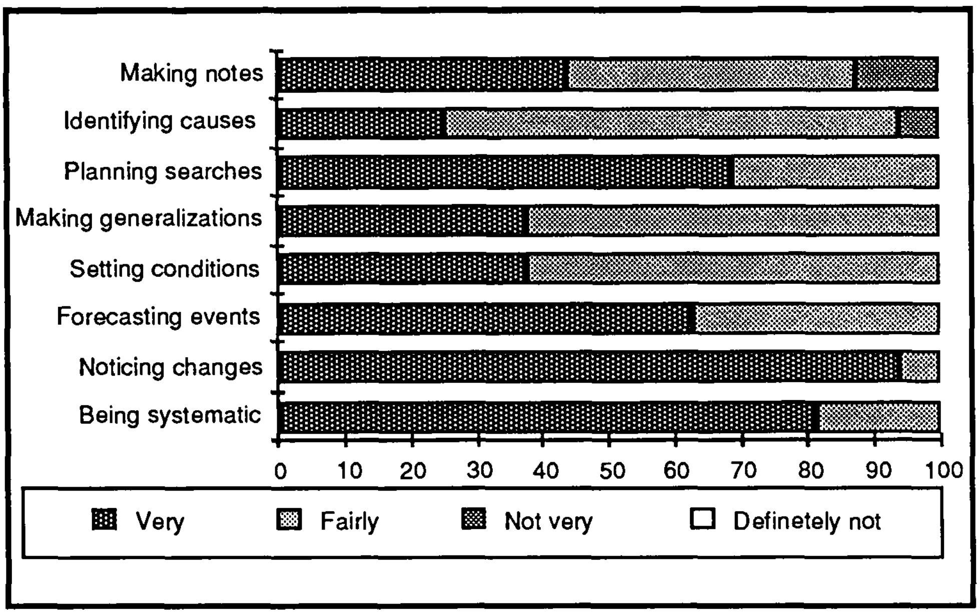


Figure 3.6: teacher's levels of agreement on the importance of certain science processes as educational aims in percentages.

When teachers are asked the extent of their agreement with these processes and a few more as educational aims, they rate highly (60% to 95% of the teachers) some processes that have less cognitive demand ('noticing changes', 'being systematic'; with the exception of 'planning searches' and 'forecasting events'), less highly (40% to 45%) processes with more cognitive demand ('making generalizations' and 'setting conditions'; except for 'making notes'), and lowly (25%) 'identifying causes', when 'very' is considered as the only element to judge the extent of the agreement (see Figure 3.6).

These results tend to suggest that:

i) Teachers expect a reasonable level of success (perhaps even too high; with Black Boxes and Balancing easier than Floating and Rolling) for the Investigation plus Goal tasks, and would use them as teaching devices for the encouragement of logical thinking and scientific processes.

ii) Teachers expect children performing very well in 'noticing', less well in 'setting conditions' and 'identifying variables', and as relatively low in 'imagining causes'.

ii) Nearly all teachers rate as important 'noticing changes' and 'being systematic', many rate highly processes concerned with searches, controlling conditions, and least highly 'identifying causes'.

Chapter 4.
DATA COLLECTION.

A considerable part of the rationale of the research has already been discussed in Chapter 3, in which the design of the tasks and their effect on the study were described in some detail.

Briefly, four tasks were devised, each in two forms: 'Investigation plus Goal' and 'Structured'. The Investigation plus Goal form in each case required the child to investigate a situation so as to find out how to reach a particular type of goal. The exact conditions under which the goal was to be achieved were not revealed, until the child was asked to say, and then demonstrate, how to reach the goal under these conditions. Because this form of task, by its nature, could not give data about how the child understood the problem (but only about what the child did) the second Structured form of task was devised. All four Structured tasks were, as far as possible, parallel in form, with sections devoted to the same set of 'processes', suitably interpreted for each task. The two forms of each task used the same apparatus and materials, and the same kinds of activities.

4.1 <u>DESIGN OF THE STUDY</u>.

The children's age, the formation of groups, what tasks children performed, how different factors intervening in the study were controlled, and more of the rationale behind the study, are described below.

- Age:

The age of the children was not taken as an experimental variable, desirable though that would have been in a larger study. Instead age was closely controlled. The tasks were designed for children ending primary school (4th year juniors). The actual ages of the children ranged from 125 to 137 months, with a median of 131 months (10 years 11 months) (see Section 6.1).

- Groups and tasks:

Two groups of children (described below) performed the whole set of tasks that is, tasks in both forms in all different content areas (Rolling, Balancing, Floating and Black Boxes). Thus, each child performed 4 tasks each in two forms (Investigation plus Goal and Structured); giving, with 24 children, 192 sessions (see Chapter 3).

Groups G1 and G2 of 12 children each were formed. One group performed the Investigation plus Goal tasks first and the Structured tasks second (see Figure 4.1). The second group performed the Structured tasks first and the Investigation plus Goal tasks second.

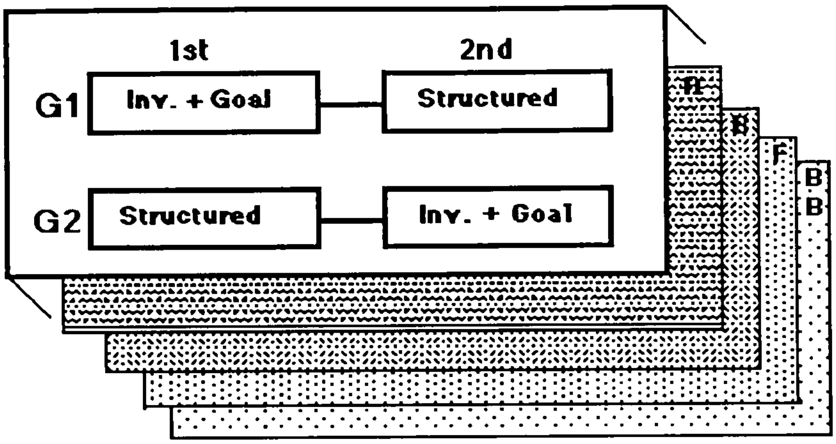


Figure 4.1: model of the design of the study.

- *Factors:*

Several design problems are resolved by having all children do all tasks in both types. This leaves to be decided the order in which tasks were to be done. It was possible to manipulate the ordering of types (Investigation plus Goal and Structured), by forming groups G1 and G2 who did the tasks with Investigation plus Goal first and Structured second, and the reverse (see Figure 4.1). It was not possible, with this number of children, systematically to vary the order of the content of the tasks. Instead this was decided by logistic considerations. It would have been better, but less practical, to have randomized the order. Table 4.1 shows the task sequence for each child.

The order of performance being reversed for the groups allows us to look for whether the order of performance makes a difference. One might think, for example, that performing the Structured tasks first could give children some understanding of the tasks, helping them to perform better (later on) in the Investigation plus Goal tasks.

The design of Figure 4.1, also allow us to know whether performing one version first or second (independently of which one), makes a difference; for example children could become more confident as time passes (the second task is performed immediately after the first).

The two different approaches or versions are performed by all children in both groups. The two types of tasks have different functions, and yield different complementary data. Thus, is not possible to compare performance on them, but it is possible to see whether performance on them correlates positively or not (Section 6.6).

Comparisons can be made of the relative difficulties of the four tasks within each type, but these comparisons are to some extent conflated with possible order effects (Section 6.3).

G1		G2	
Child	**Tasks sequences**	**Child**	**Tasks sequences**
1) Jonathan	B-R-BB-F	1) Bonnie	BB-R-F-B
2) Marie	B-BB-R-F	2) John	BB-F-R-B
3) Richard	B-F-R-BB	3) Melissa	BB-B-R-F
4) July	R-BB-B-F	4) Nowell	R-F-BB-B
5) Gavin	R-F-B-BB	5) Sally	R-B-BB-F
6) Fiona	BB-F-B-R	6) Anthony	F-B-BB-R
7) Mark	B-R-BB-F	7) Genevieve	BB-R-F-B
8)Bianca	B-BB-R-F	8) Gavin	BB-F-R-B
9) Ryan	B-F-R-BB	9) Zain	BB-B-R-F
10) Tina	R-BB-B-F	10) Andrew	R-F-BB-B
11) Edward	R-F-B-BB	11) Shaheda	R-B-BB-F
12) Jane	BB-F-B-R	12) Noah	F-B-BB-R

Table 4.1: sequences (of performance) in which children did the tasks.
B = Balancing; F = Floating; BB = Black Boxes; R = Rolling

The design is not, however, primarily intended to investigate differences between types of task, order effect, learning from one task to another, etc.. Its main function is to yield, for each child, complementary data from two forms of tasks, each 'replicated' in four content areas. The primary purpose of the study, as explained in Chapters 1, 2 and 3, is to obtain data to enable critical scrutiny of "eliciting scientific processes through tasks of a kind and content suitable for children".

The 'experimental' design of the study sees the outcome variables as measures on performance and on processes in the two types of task. Age is controlled, order effects of the two types of task were however treated as quasi-independent variables, to see if performing one type influenced performance on the other - as it turned out (Chapter 6) such effects were not evident, which was convenient for the analysis. The other main variable, content of the tasks, was treated as a replicating variable, with all contents given to all children. The whole point is to carry a critical exercise on the idea that 'processes' can usefully be elicited, rather than taking it for granted.

Chapter 6 will deal with results concerning the versions, groups and session part (first or second performance independent of version) .

4.2 COLLECTION OF DATA.

The characteristics of the sample and the organization of the data collection are now described.

- The sample:

A total of 24 children from 4th year junior classes in 2 Islington state schools performed the tasks. They were divided, as mentioned earlier, into 2 different groups (G1 and G2) of 12 children each. One school had only one 4th year junior group and the other had two. Both schools, one more than the other, show a mix of races; one is a Roman Catholic school.

Children were selected at random from three different bands of general performance (high, medium, low). To form these bands teachers in each school were asked to categorize them in such bands of ability, without any objective measure but from their own knowledge of the children. They were then selected at random in equal numbers from each category. Gender was balanced by choosing 4 girls and 4 boys in each category. Children were drawn in equal numbers from each school.

- Practical organization:

Children performed each task individually, with either the Investigation plus Goal task first and then immediately afterwards the Structured one (same content) or in the reverse order. In the majority of the cases children accomplished both forms of the task in all four contents within two weeks. They were taken out of the class to the library or to another quiet room that was available at the time. With restrictions on the use of rooms it was difficult to randomize the order of performance in relation to content for each individual (see Design of the study above).

The way tasks were presented to children has been described in Chapter 3.

In both schools I was introduced to the children and I explained the purpose of my work with them. In one school the children knew that I was a parent - my children being however in lower grades. This may have helped to get their confidence more easily, not being a complete stranger.

Chapter 5.
CONSTRUCTING DATA IN TERMS OF PROCESSES.

In later chapters results will be given about children's performance on various 'processes'. This chapter discusses how children's behaviour was categorized and how their performance was scored. That is, it concerns how these 'processes' are taken to exist.

The chapter has two purposes. One is methodological, to show how scores were defined and constructed. The other important intention is to scrutinize the way 'processes' are constructed, so as not to take them for granted. Thus the existence of 'processes' will be questioned from the beginning; as was already begun in earlier chapters. Thus, this chapter, besides containing information needed for later chapters, also contains results in the form of a discussion of the problem of arriving at useful indicators of various 'processes'. For these reasons, I shall write of 'constructing' scores, not regarding the meaning of any score as self evident.

5.1 INVESTIGATION PLUS GOAL TASKS.

A description of how data were constructed for these tasks will be given by providing a detailed account of one of the 'processes' as an example, 'searching', and then giving just the main points for the rest.

Searching is selected as an example, because of its importance in describing how extensively or completely children investigate the problem.

Data for 'processes' in the Investigation plus Goal tasks, except for 'making notes', are the records on the check lists described in Chapter 3.

5.1.1 SEARCHING.

The starting idea, as already stated, was to make possible some kind of description of the completeness of children's search of possibilities in the task. This was done by recording which variables children changed and defining groups of such patterns of variables as 'configurations'.

- *Configurations:*

One way to keep track of which combinations of variables the child tries, would be to list all possible combinations of states of all variables, were necessary treating a variable as discrete (e.g. 'gun' low or high).

It was felt that this approach needed to be modified, to take account of the fact that: all possible combinations are too many to be tried, some changes of a variable are trivial and some make a much longer change to the aspect of the problem being investigated than do others. In relation to the third fact, for example, just adding more weight at the same place on a balance stays within the same strategy for balancing, whereas moving some bricks to other places on the balance is a more substantial change of strategy.

This led to the idea of defining 'configurations' of values of variables for each task. A 'configuration' was intended to be a collection of combinations of states of variables within which the child is (or seems to be) tackling one limited aspect of the investigation. Thus all cases with 'bricks' (to follow the example of balancing) paired at equal distances on either side of the balance were taken to belong to the same configuration.

In so far as this grouping of configurations is appropriate, it makes possible a better account of the amplitude and systematism of the child's' investigation. A child who tries possibilities covering most configurations is trying a wide range of kinds of strategy. A child who tries an equal number of different combinations, but in fewer configurations, would be making a narrower search. Just counting numbers of different variable combinations tried would not give this information.

In the later analysis it will be useful to group configurations even more broadly, into groups which have something more general in common - for example all those where the same pattern of bricks is chosen on each side of the balance.

Balancing:

In this task children can change distance and weight; the former by putting bricks on different 'stripes'; and the latter by adding or removing bricks (see description in Chapter 3). The four positions of stripes (2 on each side), and the different numbers of bricks which can be put on each allow a large number of combinations. These can be reduced over by grouping similar patterns of bricks and distance as belonging to the same configuration.

The configurations chosen are tabulated in Figure 5.1.1 and illustrated in Figure 5.1.2 . In the two 'single' configurations, the same single stripe is used on either side. In the 'equivalent' case, instead of identical bricks being used, bricks are combined so as to be equivalent to another brick (e.g. 2 small bricks for a medium one). In the 'double' configuration' both stripes are used on both sides but essentially similarly. In the 'compensated' cases, (double and pairwise) attempts are made to have larger bricks closer in balance smaller ones further out. In the 'combinatorial' configurations, combinations of bricks and stripes are tried which are more general than the cases above.

WEIGHT

DISTANCE	Bound to balance		Not bound to balance	
Same	Single*/ Identical Single/ Equivalent	Double**/ Identical Double/ Equivalent Double/ Compensated	Single/ Different	Double/ Different
Different	Pairwise*/ Compensated	Combinatorial ***/appropriate	Pairwise/Not enough comp. Pairwise/ Crossed Pairwise/Not compensated	Combinatorial/ Not appropriate

Figure 5.1.1: search space, formed by configurations, for the Balancing task.
* Single and Pairwise use 2 stripes.
** Double uses 4 stripes.
*** Combinatorial use 3 stripes.

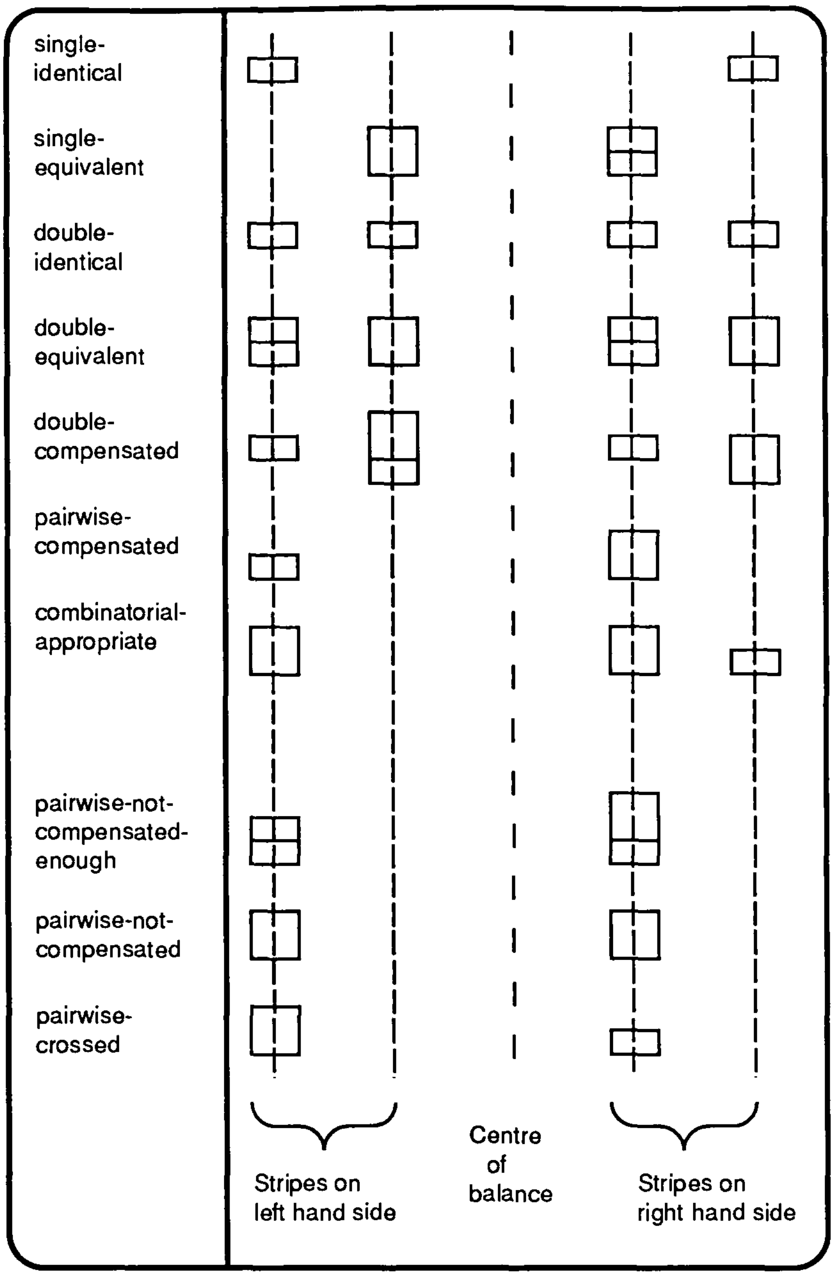


Figure 5.1.2: configurations for the Balancing task; showing sample combinations of bricks and positions of bricks for each configuration.

Figure 5.1.1 shows on which configurations the balance is bound to be level (a solution) and on which it is not bound to balance; this gives further information on performance. The four main groups in Figure 5.1.1 will be regarded as groups of configurations.

Floating:

The number of shots (ball bearings) and the solutions (beakers) tried are the two variables at children's disposal. Because of the decision not to interrupt the children's performance, so as to make the situation as spontaneous as possible, it was often not possible to know how many shots they were using. I decided, therefore that the amount of shots would be 'measured' by how far or close the straws were from the desired floating level. Thus in this case configurations are the combinations of densities (beakers) and floating levels ('far': four or more shots; 'very near': two to three shots; 'exact': the required number or one more or less; Figure 5.1.3). Those configurations that belong to the same solution (beaker) are considered as groups of configurations.

FLOATING LEVEL

DENSITIES	Far	Very near	Exact
A	+/- 4 or more required shots	+/- 2 or 3 required shots	+/- 1 required shots
B	" " " "	" " " "	" " " "
C	" " " "	" " " "	" " " "
D	" " " "	" " " "	" " " "
E	" " " "	" " " "	" " " "

Figure 5.1.3: search space, formed by configurations, for the Floating task.

Rolling:

There are two variables (velocity and direction) that children can in principle manipulate; but given the problems already described in Chapter 3, I decided to use position of 'gun' and 'trap' instead. Therefore, 'configurations' are given by all combinations of the positions of 'gun' ('up' or 'down') and 'trap' ('up' or 'down'; 'left' or 'right' and; 'upwards' (facing the top frame) or 'facing' (facing the 'firing' area) (see Figure 5.1.4 and Chapter 3). These configurations will be grouped latter (Chapter 8) as 'horizontal at top' (both gun and trap up), 'downwards' (gun up and trap down), 'upwards' (gun down and trap up) and 'horizontal at bottom' (gun down and trap down).

POSITION OF TRAP

POSITION OF GUN		Up/Left	Up/Right	Down/Left	Down/Right
Up		GuTulu	GuTuru	GuTdlu	GuTdru
		GuTulf	GuTurf	GuTdlf	GuTdrf
Down		GdTulu	GdTuru	GdTdlu	GdTdru
		GdTulf	GdTurf	GdTdlf	GdTdrf

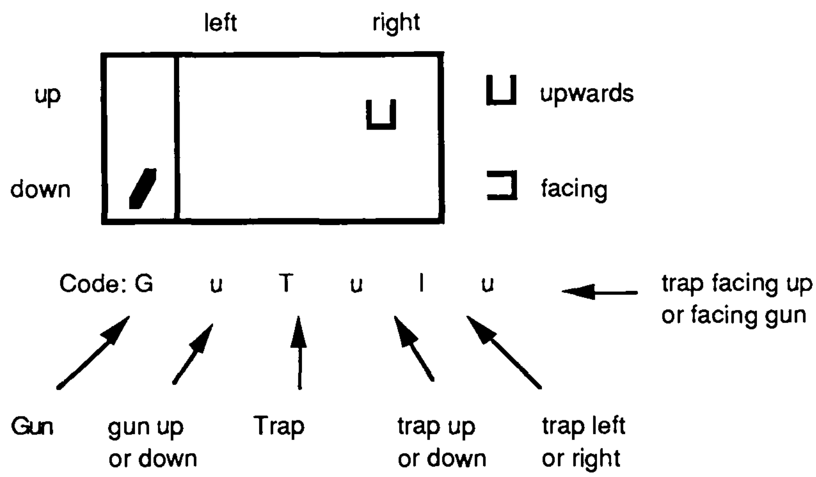


Figure 5.1.4: search space, formed by configurations, for the Rolling task (the third dimension, rotation, is built in each cell).

Black Boxes:

For this task two kinds of sets of configurations were devised; one that deals with a search within a box and the other with a search comparing boxes. In the first case, sides of the boxes ('left', 'right' and 'obstacles') and the boxes themselves (C, B, Z, I) are considered as 'variables' (Figure 5.1.5). Groups of configurations will be regarded as those belonging to a particular box.

SIDES

BOXES	Left	Right	Obstacles
C	test	test	test
B	"	"	"
Z	"	"	"
I	"	"	"

Figure 5.1.5: search space, formed by configurations, for the Black Boxes task taking boxes as entities.

In the second case, all combinations of pairs of boxes and the same sides ('left', 'right' and 'obstacles') of those pairs are considered as 'variables'. The combination of such factors gives a number of configurations, some of which can be used as tests to find out which box is which (see Figure 5.1.6). A group of configurations is those belonging to the same pair of boxes.

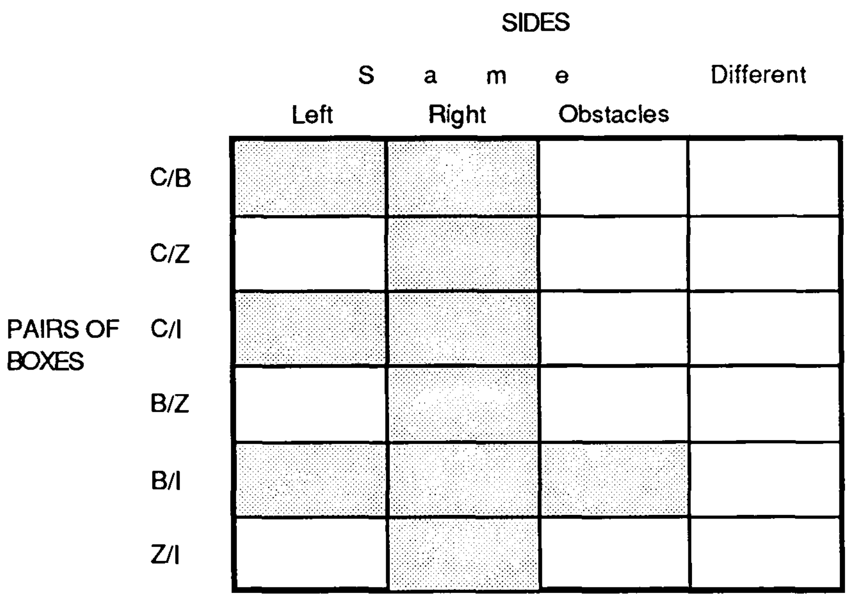


Figure 5.1.6: search space, formed by configurations, for the Black Boxes task taking boxes as analytically comparable. Comparisons that can be used as a test are shaded.

- Process of recodifying and scoring:

As explained in Chapter 3, children's behaviour was recorded on a check-list (see Appendix A). From the check lists it was possible to find the number of times each child tried something out in each configuration, or each group of configurations. It was also possible to count occasions when this followed a previous success or failure. See Appendix D (Tables 1, 2, 3, 4 and 5).

After this retrieval process, a score was constructed for each task to indicate the completeness of the child's search of the search-space of configurations - as the fraction of the different kind of attempts made (configurations) among all possible ones within a task for each child. To categorize different levels of performance, these scores were divided into three equal ranks. The individual scores, resulting from this process are used in Chapters 7, 8 and 9 to analyze 'processes', children's behaviour and relationships between the two versions of the tasks respectively.

- Discussion:

The description given of the concept of configurations for each task clearly shows that:

i) they vary in nature from task to task.

ii) this variation is due to the way variables are conceptualized and their combinations considered for each task, given the practical constraints and the intention to get valuable information in each case.

The idea of seeing how completely children investigate a problem is clearly important. Equally important for theory and practice, is the difficulty of doing so in certain cases, as shown above.

One difficulty of investigating how fully children investigate a problem was (as in floating) that actions it would have been desirable to record could be not recorded without interrupting children's work. As a result the 'search space' was to some extent artificial. Another relates to the inherent nature of the task, as in Black Boxes. Here nothing exactly corresponds to 'a change of variable'. If such a task is considered valid, then it has to be admitted that 'searching a space of variables' is not a universally valid notion. In the present case, I simply adapted it as best as I could. If a pedagogically valid task is to be used also to generate research data, its design certainly needs careful consideration, and may be considerably constrained.

5.1.2 CONTROLLING VARIABLES AND REPLICATING.

Having recorded children's attempts on the check-lists (knowing also if they succeed or failed), it was then possible to see how they changed 'variables' from one attempt to the next. Two types of changes were distinguished: when children make changes which imply going from one configuration or group of configurations to another, depending on the task and, when they make changes within a same configuration or group of configurations. The first type of changes may be more strategic and the other more tactical.

- Changes introduced when changing configuration or group of configurations:

The notation change ø, change 1, change 2, will be used to mean changes of ø (nil), 1 or 2 variables respectively.

Balancing:

Group of configurations are taken to be: 'single', 'double', 'pairwise' and 'combinatorial' (See Figures 5.1.1 and 5.1.2). To change group of configurations after success or failure, it is necessary to change the distance, though weight can be changed as well (see Table 1, Appendix H).

Floating:

In this task changing group of configurations means to change beaker. Therefore to change one variable means to change density; changing two implies changing both density and weight (see Table 4, Appendix H).

Rolling:

To change configuration in this task children can change the position of the gun, and the vertical or horizontal position of the trap, and the direction it faces This such a change may involve changing from one to four variables (see Table 6, Appendix H).

Black Boxes:

Changing one group of configurations to another in this task necessarily means changing box, and may involve changing the test applied as well (see Table 3, Appendix H).

- Changes introduced within the same configuration:

If children make no changes, but repeat the previous attempt, this can be counted as a re-trial. After a failure, it looks like a check or a second attempt; after success it looks like an attempt at replication. Scores were constructed to indicate replication.

Balancing:

Children are 'allowed', given the concept of configurations for this task, to change nothing, change 1 (either distance or weight) or change 2 (distance and weight) (see Table 1, Appendix H).

Floating:

In this case it is possible to change none (change ø: no change of beaker) or to change 1 (weight) (see Table 4, Appendix H).

Rolling:

Remaining within the same configuration in this task means to remain in the same places for gun and trap, without any changes. Under this condition, it is only possible to change 1 (either direction or velocity) or change 2 (both direction and velocity)(see Table 6, Appendix H).

Black Boxes:

To make attempts within the same configuration means, in this case, to make tests within the same box; making it possible to change none (do the same test) or change 1 (change test within the same box) (see Table 3, Appendix H).

- Process of recodifying and scoring:

Children's behaviour registered in the check-lists was recodified by comparing one attempt to the next, noting which changes were introduced after success or failure.

A score for 'controlling variables' was then devised, with different scores given to different kinds of attempt. The scheme gave the highest score when only one change was introduced (whether after success or failure), less when two changes were made after success and the lowest when two changes were made after a failure (see Tables 2, 5 and 7, Appendix H). Black Boxes was excluded since it did have not success or failure as an outcome. A separated score was constructed for 'Replicating,' being the number of attempts making no changes after success, divided by the total number of attempts made after success.

-Discussion:

The importance of a process such as 'controlling variables' is widely admitted. What the previous account shows is the considerable complexity of measuring any such process, in a comparable way in different tasks. However, the account also suggests that some progress can be made provided that a careful analysis of the task is undertaken, in particular using some such notion as the 'configurations' applied here.

Replication, though of admitted importance in science, less often features in lists of 'scientific processes'. Perhaps it lacks glamour. However, it appears here as an example of a process which has an a-priori basis, appears natural with the context of the task, and can with some effort be given a suitable measure. It is important to note that it can only be elicited at all when the child is left free to decide what to do, as in the Investigation plus goal tasks. Also, the construction of a score points to a subtle but important fact: replication cannot be measured just by repetition - it matters whether the repetition follows an attempt which in some sense succeeds or fails.

5.1.3 MAKING NOTES.

As explained in Chapter 3, children were given a blank sheet of paper with the advice that it could be used if they needed it. Children's use of the paper was classified as 'not making notes', 'making irrelevant notes' and 'making relevant notes'.

- Discussion:

'Making notes' does not, so far as I am aware, appear on any list of 'scientific' or cognitive processes relevant to primary school science. But it can be argued that in any situation other than the simplest, human beings have great need to rely on forms of 'external memory'. The limits of short term memory are well known. thus the process of recording (ideally economically and vividly) what has been seen or found, is arguably vital to science education. That it can be represented among the 'processes' discussed here owes something to the accident of the experimental procedure.

5.1.4 WHAT CAN BE CONCLUDED FROM THE CONSTRUCTION PROCESS?

The main conclusions which seem to me to be suggested by this process of construction are:

i) the intention to implement the same scheme of analysis for the same intended behaviour or 'process' finds obstacles it is difficult to avoid.

ii) the implementation of the same scheme of analysis for different tasks is restricted by special features of each task, and it may require special analytical devices such as 'configurations'.

iii) the meaning of the classification system for children's performance (and thus the meaning of different levels of performance), mostly depends on the way categories are decided upon. In some cases, an arbitrary score is constructed (e.g. search), when performance can only be rated in terms of norms. In others (making notes), an evaluation of relevant material can generate a more criterion referenced score.

iv) interesting processes are not necessarily easy to score.

v) processes which are easy to score are not necessarily interesting.

vi) there may be processes not on published lists which are worth attending to.

5.2 STRUCTURED TASKS.

The construction of 'processes' in this version of the tasks, is different from that described for the Investigation plus Goal tasks, because of the different methods used to elicit children's behaviour. In the Structured tasks questions attempt to elicit consistent behaviours from pre-determined 'processes'.

The first step in codifying responses was to categorize all the responses qualitatively. Appendix E gives the categories used for each question in each task. These categories were then used in two ways. First, they formed the basis for arriving at numerical scores for each response. The treatment of these scores is discussed below. Secondly, they were used in interpreting results (reported in Chapter 7 and 8) from correlations between different processes in the same task and the same process in different tasks.

The description and analysis of how scores for possible 'processes' were constructed in this version, will use two 'processes' as examples ('noticing' and "what-if' reasoning') and will include a description of the classification process, the validation process of the categories, the adding up of consistent behaviours within the same 'process', the scoring process and the problems found. It is thought that these two 'processes' are representative of the procedure followed and the problems found for all 'processes'.

5.2.1 NOTICING.

As described in Chapter 3, the 'process' termed 'noticing' involves perceiving or noticing changes or differences, using apparatus or materials.

- *Categorization and validation processes:*

The process of construction of data starts from the decision to attempt to elicit certain behaviours in a particular way. What it is desired to elicit and means adopted to do do can be seen in Appendix B (the instruments).

The next step was to examine children's answers and to decide, a posteriori, on categories of response. To achieve reliability in allocating responses, I and one other colleague (different for each task) made allocations independently. When differences were found, they were discussed until an agreed solution was reached. The final categories can be found in Appendix E.

It was expected that children's knowledge and expectations would affect their responses for Rolling, but the cognitive aspect of noticing in Balancing ('give reasons' as a discriminating factor) was not expected or defined before hand.

- *Scoring process:*

Categories were scored according to the level of the answers, keeping as far as possible to three levels of performance, scored 2, 1 and Ø.

- *Discussion:*

Criteria to decide levels of response could not in most cases be decided beforehand. There were generally unanticipated but evidently relevant features of responses to be taken into account. As a result criteria for different tasks were somewhat different.

For example, giving a reason to what was noticed (Balancing); noticing floating levels with detail, noticing floating levels in relation to weight used, and noticing both effects of different masses and densities at the same time (Floating); noticing the correct difference (Black Boxes) and; noticing plausible paths (Rolling), were features used in the different tasks to discriminate levels of performance.

It follows that a difficulty must be recognized in attributing the same 'process' to responses in different contexts. The attempt here was to keep the differences as small as possible so that the 'process' retained as much coherence as possible. This of course is a general difficulty, with the concept of 'process'.

- *Adding up process:*

The next step, given that there were similar subtasks on each process in each task, was to see whether performance on these was consistent. The further question of whether consistent behaviours within the 'same' process across different contents are present or not, is addressed in Chapter 7.

The first indication of consistency was taken from correlation tables (see Tables 1, 2, 3 and 4, Appendix F). On such tables, one per task, can be seen the relationships between subtasks belonging to the same intended 'process'; as well as their relationships with all the other subtasks belonging to other intended 'processes'. These correlation tables indicate satisfactorily large correlations for a considerable number of pairs of subtasks, from which one can get some indication of whether subtasks might usefully have their scores combined.However, these Pearson correlations have two defects as a source of such judgements. Firstly, the scores on which they are based are on a coarse scale (0,1 or 0,1,2) which could almost be seen as nominal, for which the validity of this correlation coefficient is dubious. Secondly, and more important in the present case, two different reasons for a small correlation are not distinguished: either a lack of association, or a distribution such that essentially no correlation can appear. For these reasons it was felt better to look at the association directly through contingency tables.

What followed was to look at at 2x2 contingency tables and to decide whether it would make sense, from the shape of the relationship and from the actual answers, to add scores together or not on subtasks within a process. Such tables can be seen in Table 5.2.1, after collapsing categories to facilitate looking for the shape of the relationships, not just relying on a number (either chi square or a correlation coefficient). It will be explained how decisions were taken from Table 5.2.1 in order to add up subtasks belonging to the same intended 'process'. In what follows subtasks within noticing are labelled N1, N2, etc. .

Balancing:

In the way the task was set up, 'noticing' could have only involved seeing whether the balance was level or not - a mainly perceptual matter not evidently requiring reasoning (see criteria for the tasks in Chapter 3). For this reason, the giving of a reason for the balance being level or not was included in the score for 'noticing'. I realize that this decision is debatable, and could be seen as destroying the coherence of the concept of 'noticing'. On the other hand, if observation depends, as most agree it does (Hodson R., 1986a &

1986b), on what we understand or expect, this inclusion looks less unreasonable.

Both subtasks required giving reasons: nobody who failed in giving reasons for N1 (balance tilted) did it for N2 (balance level); exhibiting a positive relationship. The difficulty of N2 is that children come from N1 using the same weights and stripes, but now the balance is level; which appears difficult to give a reason for if children do not understand the relationship between weight and distance. In this case, subtasks were combined, naming the 'process' : 'noticing' total (NT).

Black Boxes:

When the difference in distance to be perceived is fairly noticeable (N2) children score well, but when the difference is smaller noticing becomes more difficult (N1) and the subtask more discriminating. In fact N2 adds very little information to N1, merely that some children are not quite capable of noticing a fair distinction. Economy suggest adding them together.

Balancing N1

N2	1	17	11	6
	2	7	0	7
			11	13
			1	2

Black Boxes N1

N2	0	2	2	0
	1	22	8	14
			10	14
			0	1

Floating N1

N2	1	2	2	0
	2	22	10	12
			12	12
			1	2

N1

N3	0	6	2	4
	1	18	10	8
			12	12
			1	2

N2

N3	0	6	1	5
	1	18	1	17
			2	22
			1	2

Rolling N1

N2	0	18	12	6
	2	6	1	5
			13	11
			0	2

N1

N3	0	15	8	7
	2	9	5	4
			13	11
			0	2

N2

N3	0	15	12	3
	2	9	6	3
			18	6
			0	2

Code: N1 = Noticing subtask 1; N2 = Noticing subtask 2; N3 = Noticing subtask 3

Table 5.2.1: contingency tables for 'noticing' within a task.

Floating:

N3 (notice at the same time the effect of different masses and weights) looks as if it were not comparable to N1 and N2 due to the low level of performance; in fact nobody gets the highest mark in N3. N1 shows the best power of discrimination, with N2 showing very limited discrimination. N1 and N2 are added together (forming N12) to keep the small amount of information contained in N2; but N3 looks too different to be added to N1 and N2.

Rolling:

The strongest positive relationship is between N1 and N2. N3 looks different from both because in scoring N3 it was decided to include 'straight' paths as plausible, unlike N1 and N2 (where they are parabolic). Thus, N1 and N2 are added together giving N12, leaving N3 on its own.

The 'processes' formed by this adding up process (NT for Balancing, NT for Black Boxes, N12 and N3 for Floating and N12 and N3 for Rolling are used in Chapter 7 to see whether a 'process' (consistent behaviour within individuals and variability between individuals) across different contents or tasks is present or not.

- Discussion:

The essential criterion to decide whether scores on two or more subtasks should be added together or not is a balance between signs of a positive relationship (with a reasonable distribution on both sides) and qualitative evidence from the responses that behaviour on such subtasks have similar meaning (e.g. reason why N3 was not added in Rolling). Another criterion is economy grounds: not distinguishing scores when there is little reason to do so - reason why N1 and N2 where added together in Balancing. This procedure will tend to exclude as 'processes' behaviours that show consistency within individuals but little variability and so no possibility of seeing a correlation. Thus processes too easy or too hard tend to be excluded; the question remains whether those behaviours are or are not part of a bigger dimension (process?), part of some development.

5.2.2 WHAT-IF REASONING.

Here it is intended to elicit, as described in Chapter 3, the ability to forecast (in a loose sense) an outcome, given certain conditions; that is, to say 'what would happen if...' . As with noticing, the performance on different subtasks, were examined to decide whether or not to combine them.

- Categorization and validation processes:

The procedure was the same as described for noticing. The instruments to elicit 'what-if reasoning behaviour can be seen in Appendix B.

- Scoring process:

Essentially the same procedure as with noticing was followed. The main difference is that in this case children were required to justify their prediction (except for Black Boxes), so that in order to get the highest score they had both to make a correct prediction and to give a sound justification.

- Adding up process:

Table 5.2.2 shows the 2x2 tables for 'what-if' reasoning after collapsing.

Balancing:

Predicting the weight needed to get the balance level seems to relate strongly to predicting the required distance to get the balance level. Thus these scores were combined.

Balancing		WR12	
WR34		0	2
0	12	11	1
2	12	3	9
		14	10

Floating		WR12	
WR34		0	2
0	10	7	3
2	14	0	14
		7	17

Rolling		WR12	
WR34		0	2
0	18	13	5
2	6	3	3
		16	8

Black Boxes		HR12	
HR34		0	2
0	7	6	1
2	17	4	13
		10	14

Code: WR12 = 'what-if' reasoning subtasks 1 & 2
WR34 = 'what-if' reasoning subtasks 3 & 4
HR12 = hypothetical reasoning subtasks 1 & 2
HR34 = hypothetical reasoning subtasks 3 & 4

Table 5.2.2: contingency tables for 'what-if' reasoning.

Floating:

Again the scores show a strong positive relationship; that is, successful predictions of the highest and lowest floating levels of a straw with constant weight are consistent. Subtasks were added together, resulting in WRT.

Rolling:

The principle of economy was applied here, since the distribution is not very evenly divided, giving a weak but positive relationship. The process scores were combined.

Black Boxes:

Children's ability to make a discrimination and to point out a lack of discrimination seems to relate, as reflected by a strong and positive relationship. They were thus combined, resulting in HRT.

- Discussion:

As with noticing, the empirical data both provides a basis for a potential critique of the 'process' in question, and is necessary in arriving at reasonable 'process' scores. Had there been only one subtask for each process, such data would not have been available. In the present case, it seems that 'what-if' reasoning emerges as a viable 'process' in most tasks.

5.2.3 REMAINING 'PROCESSES'.

Having described in detail the construction of 'processes' for 'noticing' and 'what-if reasoning', the way subtasks (belonging to a same process) were combined for the remaining processes in the Structured version of the tasks will be simply reported.

Processes like 'identifying variables', 'what-if reasoning' and 'understanding' (Rolling) include identification, prediction or expectation and their justification for each answer. In these cases both answers are taken together in terms of the scoring system (e.g. identification of a variable + a plausible justification = highest mark).

Table 5.2.3 shows how the remaining subtasks ('understanding', 'identifying variables', 'making generalizations' and 'imagining causes') were combined. In all cases but one there is in the end one score for each 'process' on each task - whether because scores are combined or because only one question was originally asked.

For Floating, 'imagining causes', the score for C2 is kept apart from C1+C3, because C1 and C3 both concerns fluids whereas C2 concerns weight. In Black Boxes, 'identifying variables', responses to two questions are ignored. These questions, on consideration, did not reflect 'identifying variables' in any meaningful sense.

Content / 'Processes'	Balancing B	Floating F	Rolling R	Black Boxes BB
Understanding (U)	U	U	U1 + U2 = UT	U
Identifying variables (V)	V1 + V2 = VT	V1 + V2 = VT	V1 + V2 = VT	V1*, V2, V3
Making generalizations (G)	G1 + G2 = GT	G1 + G2 = GT	G1 + G2 = GT	G1 + G2 = GT
Imagining causes (C)	C1 + C2 = CT	C1 + C3 = C13 C2	C1 + C2 = CT	C

Code: digits 1, 2 and 3 indicate the subtask forming the task or process; T = total (when all subtasks within a task were added up).

Table 5.2.3: description of how subtasks were added up in each 'process' and task.
*V2 and V3 were ignored here: given their phrasing responses could not be counted as 'identifying variables'.

Chapter 6.
TASKS AS PRACTICAL ACTIVITIES.

In this Chapter levels of performance on both versions of the tasks will be described as well as some general characteristics belonging to children and tasks; in Investigation plus Goal tasks in relation to how well children predict and practically solve the problem posed, and in Structured in relation to their overall performance on all sections of the task.

It will start with general features of the tasks and of the children who performed them and will continue with levels of performance and difficulty of the tasks, consistency of children's performance across tasks, any relationships between versions, and possible learning effects.

6.1 WHAT AGE WERE THE PUPILS?

The tasks, as explained in Chapter 3, were designed for children ending primary school; that is, for 4th year juniors.

Figure 6.1 shows that their ages have a median around 131 months (almost 11 years old) with the values showing a distribution ranging from 125 to 137 months. These ages look typical of children starting as 4th year juniors, at not less than 10 years but below 11.

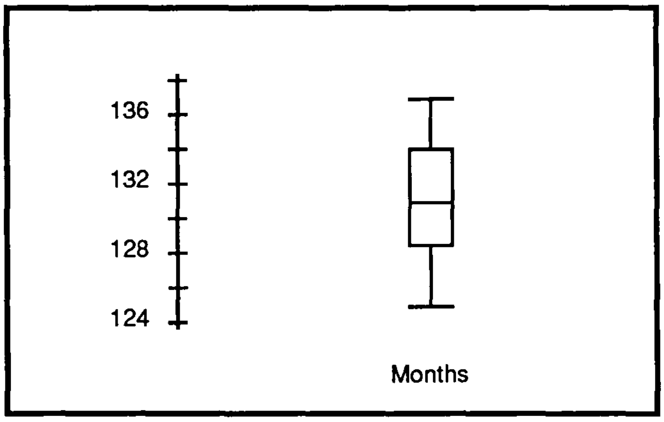


Figure 6.1: distribution of children's age in months.

6.2 HOW LONG DO THEY TAKE?

It can be seen from Figure 6.2 that the Investigation plus Goal versions of the tasks last on average half as long as the Structured versions; with the time for the Structured versions varying more than for the

Investigation plus Goal versions. Twenty minutes on average, as for the Investigation plus Goal tasks, seems a reasonable time to perform an activity inside the classroom. As expected, Structured tasks lasted longer due to the intention to elicit six different 'processes'.

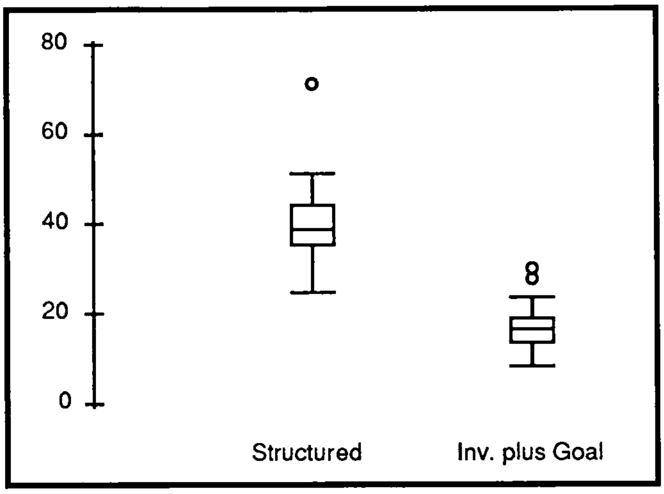


Figure 6.2: distribution of average time taken in minutes per version.

In Figure 6.3 it can be seen that, of the Investigation plus Goal tasks, Black Boxes takes almost half the time the others take (with some few exceptions), possibly because the manipulations are not as lengthy or as accurate as in the other tasks. The other tasks last more or less the same time; with a few children in Floating taking outstandingly longer, and one child in Rolling taking an outstandingly shorter time.

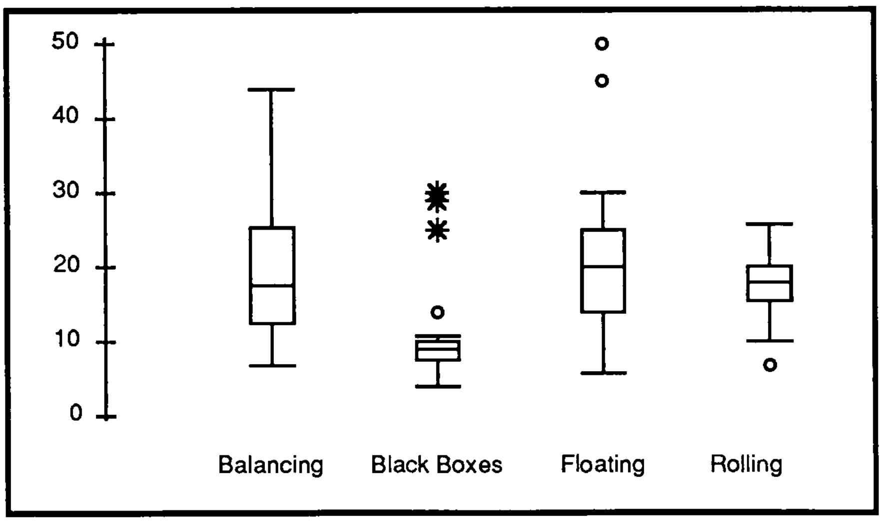


Figure 6.3: distribution of time taken in minutes for the Investigation plus Goal tasks.

Regarding the Structured tasks, Figure 6.4 shows that they all took rather similar times.

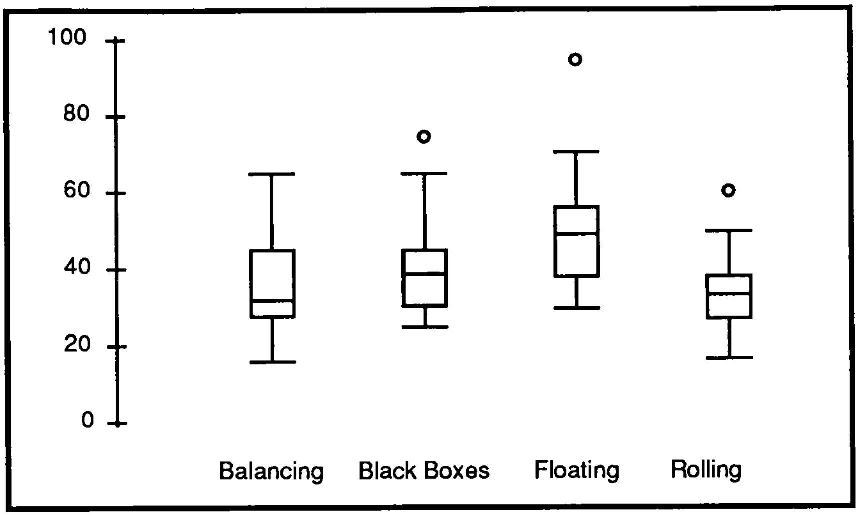


Figure 6.4: distribution of time taken in minutes for the Structured tasks.

In both versions Floating is the longest task, perhaps because of the necessary manipulations of ball bearings.

6.3 <u>CAN CHILDREN SUCCEED?</u>

In order to judge the extent of success in the Investigation plus Goal tasks, scores were given as follows: performances were given two points if children rightly predicted the solution of the problem, one point if the answer given was near the solution, and zero if it was far from it. For the Structured tasks, since total scores varied from 12 to 14, performances on these tasks were calculated as the fraction of the total possible score (adding all points achieved on each process and dividing by the total possible). To make possible a rough measure of success for this version, the following argument can be made: given the fact that the tasks have at least two questions per section ('process')(except for 'Understanding', with only one) it can be proposed that a 50% level of success corresponds to answering at least one question per section at the highest level (performances are scored Ø, 1 or 2 for each question or subprocess and then recoded Ø, 1 or 2 for the process). This appears on the face of it, to be an acceptable level of success.

Figure 6.5 shows the results for the Investigation plus Goal tasks. They show that a good fraction of children were able to perform all tasks. A reasonable proportion of children in all tasks get the highest score for a good solution of the problem, and in each case there are more children in all performing at a 'medium' or 'high' level than at 'low' level. The hardest task

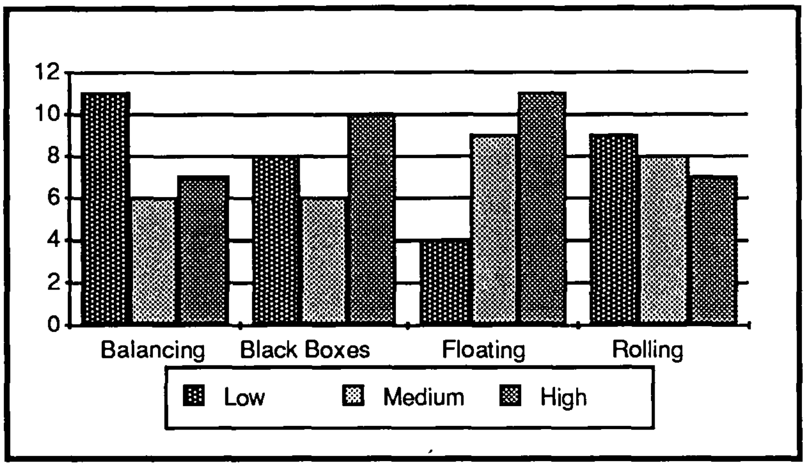


Figure 6.5: levels of success per task in the Investigation plus Goal version.

was Balancing, but it still has 7 children (29%) doing well and a little more than half doing moderately or well. Children get all three types of scores - 2, 1 and zero - in all the tasks; meaning that the tasks do have some power of discriminating primary school children's behaviour, at least with this sample.

Levels of performance in the Structured version of the tasks are shown in Figure 6.6. Using the rough measure of success already described, it can be seen in Figure 6.6 that the majority of children perform above the 50% level of success in Black Boxes and Floating, that few children reach this level or higher in Balancing, and that very few do so in Rolling. The power of discrimination of these tasks is similar and high for Balancing and Floating; with Black Boxes lower and Rolling showing a rather restricted power of discrimination (0 to 60%).

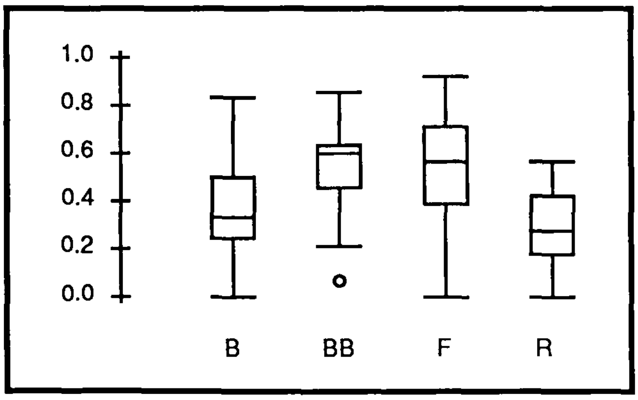


Figure 6.6: distribution of levels of performance on the Structured tasks.

These results show that in all cases but two (Balancing and Rolling in the Structured version), the majority of children perform at intermediate and high levels of success (in the way success has been defined). Overall, a good proportion of children are able to succeed, to a reasonable extent, on both versions of the tasks.

6.4 ARE SOME TASKS MORE DIFFICULT THAN OTHERS?

From the previous section it is possible to see that some tasks appear more difficult than others. In what follows levels of difficulty of the tasks will be described by the levels of performance on each task: looking at the median and distribution for individual ratio scores in the Structured tasks, and at the average score for each task in the case of Investigation plus Goal.

Figures 6.6 (Structured) and 6.7 (Investigation plus Goal) show that Balancing and Rolling proved to be the most difficult tasks in both versions, with Black Boxes and Floating easier also in both versions. The ordering of difficulty is not exactly the same, suggesting that the relative level of difficulty among tasks may change with the version. The particular difficulties of each task will be apparent in Chapters 7 and 8. However, it will be shown in Section 6.7 that different versions of the tasks do not make a difference to children's performance, except for Rolling.

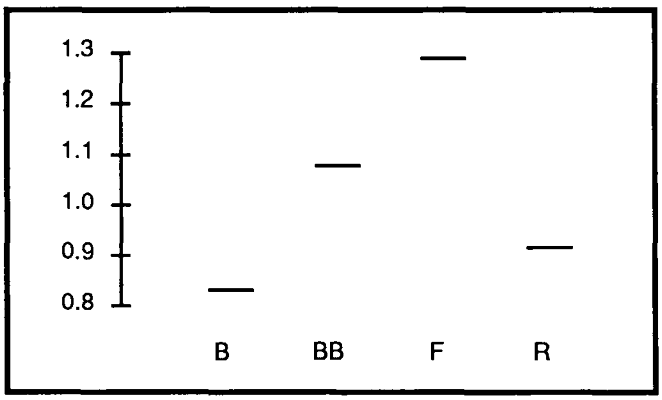


Figure 6.7: average performances on the Investigation plus Goal tasks.

Levels of performance and their distribution on the Structured version of the tasks can be seen in Figures 6.6. Balancing and Rolling show medians of around 30%, and Black Boxes and Floating of around 60%, of the maximum; suggesting a considerable difference in performance between the two pairs of tasks. Variability in performance changes from task to task, with Balancing and Floating varying most. These results suggest that children's abilities are more stretched by Balancing and Floating. These tasks are also those which are most alike in terms of the 'cognitive' abilities elicited: 'making generalizations', 'identifying variables' and 'what if reasoning' (Chapter 7).

Figure 6.7 shows the average scores for each Investigation plus Goal task. It can be seen that Balancing and Rolling have an average score of around 0.9 (maximum = 2); with Black Boxes and Floating each having average scores just above 1. These results suggest not such a big difference between the easy and the difficult tasks as in the case of the Structured version; although a caution should be given in relation to the restrictions imposed on variability given the criteria used to score the Investigation plus Goal version.

6.5 HOW CONSISTENT ARE LEVELS OF PERFORMANCE ACROSS TASKS?

The relationships between levels of performance (using the original raw scores) for the Structured versions of the tasks are shown in Table 6.1. The correlations are all positive, and most are substantial, suggesting that children perform rather consistently across tasks. A possible exception is the relationship between Black Boxes and Rolling; it will be argued later in Chapter 7 that these tasks have special features.

	B	BB	F	R
B	1.000			
BB	0.677	1.000		
F	0.709	0.618	1.000	
R	0.649	0.332	0.492	1.000

Table 6.1: Pearson product-moment correlations between scores on the Structured tasks.

In the Investigation plus Goal tasks children's behaviour also relates positively between all combinations of tasks, as can be seen from Table 6.2 (here and in Table 6.3 the scores 'low' and 'medium' have been collapsed, giving just two levels of performance.) Floating and Black Boxes have the strongest relationship, perhaps because these are the two easier tasks in terms of the smaller size of the search space. However less reliance can be placed on correlations looked at in this way, given the restrictions in variation of the scores.

			B		BB		F	
			0	1	0	1	0	1
BB	0	14	11	3				
	1	10	6	4				
F	0	13	11	2	10	3		
	1	11	6	5	4	7		
R	0	17	14	3	11	6	10	7
	1	7	3	4	3	4	3	4
			17	7	14	10	13	11

Table 6.2: contingency tables for levels of performance on the Investigation plus Goal tasks, per task.

6.6 DO LEVELS OF SUCCESS BETWEEN BOTH KIND OF TASKS RELATE?

To look at this question scores on both Investigation plus Goal and Structured tasks were dichotomized (as explained before) as in Table 6.3. Scores for the Structured version were recoded from the individual scores by forming three ranks and then collapsing 'low' and 'medium' to give two categories (zero and 1).

		S/B 0	S/B 1		S/BB 0	S/BB 1
IG/B	0	9	2	IG/BB	3	5
	1	4	9		1	15

		S/F 0	S/F 1		S/R 0	S/R 1
IG/F	0	2	2	IG/R	9	0
	1	4	16		6	9

Table 6.3: contingency tables for levels of success on both versions, per task.

All the relationships are positive and in every case 75% of the children perform in the same way on both ; only 25% perform differently. Thus overall children's performance was rather consistent as between the two versions.

6.7 ARE THERE SIGNS OF LEARNING EFFECTS BY PERFORMING DIFFERENT KINDS OF TASK?

As it was explained in Chapter 4, one group of children performed the Investigation plus Goal version of the tasks first, followed by the Structured one, with the other group performing the other way round. The groups performed both versions of the tasks. There was one factor that could not be controlled; the order in which the four tasks (contents) were performed by each child. Therefore the source of variance is focused on the effects of versions (Investigation plus Goal/Structured), session part (first/second) and order of performance of the two versions (Group 1/Group 2).

To determine whether the order in which children carried out the two versions of a task affected their performance, Lindquist (1953) Type II analyses of variance with session part and task version as within-subjects factors were conducted on the scores for each task separately - in this approach the order effect is a between-subjects component of the interaction between the two within-subjects factors.

Scores for the Structured tasks are based on individual scoring and the original values of the the Investigation plus Goal tasks were transformed (multiplied by a factor) to make them comparable to the Structured ones in absolute value.

Results of such analyses are shown in Tables 6.4a to 6.4d. For none of the tasks was the interaction significant (all F's < 2). There was no evidence that children did better on the second version of a task (all F's < 2). Only on one task, Rolling, did the scores on the two versions differ significantly, $F(1,22) = 4.76$, $p < .05$. Children scored higher on the Investigation plus Goal version (mean = 6.41, s.d. = 5.8) than the Structured version (mean = 4.33, s.d. = 2.18). Means and standard deviations for the versions of the other tasks are presented in Appendix G.

These results will simplify the discussion in Chapter 8 of evidence from the two different versions of the tasks.

	SS	df	MS	F	p
BETWEEN					
AB (G1 & G2)	1.02	1	1.02	0.045	<1
Error (b)	542.96	22	22.62		
WITHIN					
A (version)	3.52	1	3.52	0.379	<1
B (session part)	6.02	1	6.02	0.649	<1
Error (w)	203.96	22	9.27		

Table 6.4a: analysis of variance for children's performance on the Balancing task.

	SS	df	MS	F	p
BETWEEN					
AB (G1 & G2)	35.02	1	35.02	1.28	<1
Error (b)	598.3	22	27.19		
WITHIN					
A (version)	0.52	1	0.52	0.01	<1
B (session part)	38.52	1	38.52	2.48	<1
Error (w)	340.46	22	15.47		

Table 6.4b: analysis of variance for children's performance on the Black Boxes task.

	SS	df	MS	F	p
BETWEEN					
AB (G1 & G2)	1.02	1	1.02	0.035	<1
Error (b)	630.96	22	28.68		
WITHIN					
A (version)	31.68	1	31.68	3.11	<1
B (session part)	1.02	1	1.02	0.1	<1
Error (w)	223.8	22	10.17		

Table 6.4c: analysis of variance for children's performance on the Floating task.

	SS	df	MS	F	p
BETWEEN					
AB (G1 & G2)	1.33	1	1.33	0.04	<1
Error (b)	634.92	22	28.86		
WITHIN					
A (version)	52.08	1	52.08	4.76	<0.05
B (session part)	8.33	1	8.33	0.762	<1
Error (w)	240.59	22	10.93		

Table 6.4d: analysis of variance for children's performance on the Rolling task.

6.8 HOW DO TEACHERS EXPECTATIONS RELATE TO CHILDREN'S PERFORMANCE?

In Chapter 3 results were given in relation to teacher's reactions and expectations in relation to the tasks as pedagogical devices and children's performance on them, and in relation to the extent of their agreement with 'scientific' processes as educational objectives and their expectations of children's performance on them. Having dealt with children's performance on the tasks in this chapter, a comparison will be made with teachers' responses in order to see to what extent the tasks can be regarded as both feasible and corresponding to teachers' expectations.

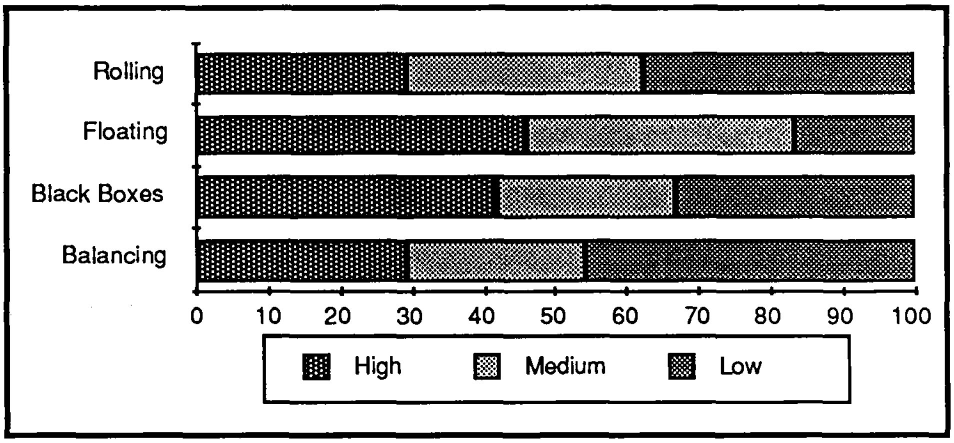


Figure 6.8: children's performance on the Investigation plus Goal tasks in percentages.

Teachers expect children performance on the Investigation plus Goal tasks to be between 34% to 48% (Figure 3.1) if 'do well' is used to measure the extent of their expectations. These figures roughly correspond to actual children's performance: 29% to 46% if 'high' is the measure of success (Figure 6.8). Although there seems to be a broad agreement, comparison between both sets of figures is not without difficulties. While teachers expect children's lowest performance ('not do well') to be around 20% for all tasks, in reality children's lowest performance ('low') vary between 16% to 46%; thus, taking the highest children's performance and the better expectations teachers show as means of comparison, only gives a partial picture of what really happens. In relation to the relative difficulty of the tasks, it is judged by the same measurements as before (with the same problems of comparability), teachers expect Black Boxes and Balancing as the easier ones and Floating and Rolling as the most difficult ones. Children's actual performance reveals that Floating and Black Boxes are the easier ones and Rolling and Balancing as the difficult ones. There seems to be, then, a rough agreement in relation to the relative difficulty of Black Boxes and Rolling, but a misjudgement in relation to Floating and Balancing; maybe by overestimating the difficulty of manual dexterity in Floating and underestimating the cognitive difficulty of Balancing.

Additionally, if an average value for children's performance is computed for all four tasks (based on Figure 6.7), a value of 1.03 out of 2 (average = 'medium') corresponds to the majority of teachers rating the tasks as 'fairly' difficult (Figure 3.4). From these results, there seems to be a certain level of correspondence between what is expected from children and their actual performance, and a willingness from the majority of the teachers in using several or all the tasks - mainly because of the encouragement given to use logical thinking or scientific processes (Figures 3.2 and 3.3).

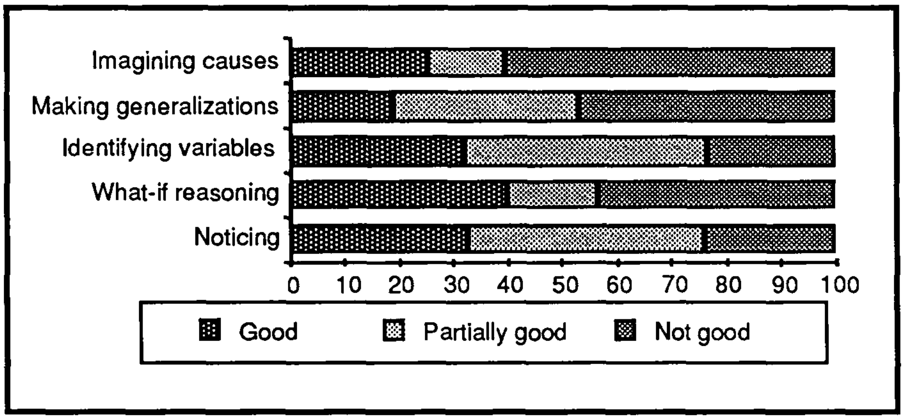


Figure 6.9: children's levels of performance on certain processes in percentages.

Comparisons between teacher's expectations and children's performance on certain 'scientific' processes seem more difficult, due to the lack of objective guidelines of success (contrary to what happen with the investigation tasks) and definition (teachers were not given a written statement of the meaning of each process, although an oral explanation was offered for 'setting conditions'). Bearing this in mind, teachers correctly forecast the easiest and the most difficult processes, this is 'noticing' and 'imagining causes' respectively (see Figures 3.5 and 6.9) if 'well' plus 'in between' and 'good' plus 'partially good' are taken into account to judge their level of difficulty. In the cases of 'identifying variables' and 'making generalizations, there is a rough agreement of their relative difficulty. The only case where seems to be a misjudgement by teachers is in 'predicting' ('setting conditions', 'what-if reasoning'). Although there seems to be a broad agreement in relation to the relative difficulty of processes, results show that teachers tend to overestimate children's performance. Overall, however, teachers predictions seem to be more in line with children' performance that the study by Black, Harlen and Orgee (1984) might lead one to expect.

It can be concluded that:

i) there seems to be some evidence that the Investigation plus Goal tasks look like potential pedagogic devices,

ii) certain 'scientific' processes are considered by teachers as valuable and achievable educational objectives, although some more than others ('imagining causes' rated more difficult than 'noticing'),

in the face of teacher's judgements and empirical evidence derived from performance on the tasks.

Chapter 7.
HOW MUCH DO PROCESSES DEPEND ON TASKS?

7.1 STRUCTURED VERSION OF THE TASKS.

7.1.1 WHAT IS THE NATURE AND EXTENT OF PROCESSES?

Inspection of the correlations between scores on subtasks (Appendix F) suggests that there exist at least some substantial correlations between processes within and across tasks. However, as mentioned in Chapter 5, Pearson correlations are liable to be misleading for these data. In particular, a method is needed which can distinguish between low correlations where there could have been a relationship but there is not, and low correlations arising because nearly all children do very well or badly on one or both tasks. Thus, as in Chapter 5, 2x2 contingency tables will be used to examine possible relationships. In forming such tables, scores have been collapsed in such a way as to maintain, as far as possible, comparability between levels of performance represented by the same score.

With 24 children, such tables can provide at best weak statistical evidence. The approach will be to discuss qualitative similarities and differences in the nature of the responses or behaviours counted as falling under each 'process', and to attempt to account for any patterns seen in the relationships of performances in these terms.

A simple approach to looking at the tables is taken from log-linear analysis. Given the table

a	b
c	d

the logarithm $\ln \alpha$ of the crossproduct ratio $\alpha = ad/bc$ indicates the sign of the relationship. The standard deviation δ of $\ln \alpha$ is asymptotically equal to $\sqrt{(1/a + 1/b + 1/c + 1/d)}$. This gives some indication of the confidence that can be placed in a given value of $\ln \alpha$. A value which does not exceed zero by one standard deviation is difficult to take seriously. A value which exceeds zero by two standard deviations probably deserves some confidence. As can easily be seen, any relationship would have to be rather strong to meet such criteria. In no instance can the standard deviation of $\ln \alpha$ be less than 0.8, because of the modest number of cases.

There is one case for which $\ln \alpha$ is less appropriate, which is when one cell is nearly empty. This usually corresponds to few children doing well on a harder task and less well on an easier one. An appropriate indicator here may be d, defined as

d = 1 - observed frequency in 'empty' cell/expected frequency in 'empty' cell.

The question to be addressed is whether children's performances in the different tasks are appropriately given one process label, that is whether their performance is best accounted for as versions of one process, or by special features of the tasks.

The way this question is answered is essential for the understanding of children's performance as reflecting 'scientific processes', for clarifying what such processes are and how to assess them, on the understanding that the above mentioned 'processes' were part of tasks intended to be reasonable classroom tasks; something to be discussed in Chapter 9.

7.1.1.1 Noticing.

It is important to draw attention to differences in what was counted as 'noticing' and to differences in dichotomizing the scoring between the tasks.

Floating (12) and (3):

In subtasks 1 and 2 (which are added together as: (12)) pupils had to notice effects of physical quantities - weight and density - one at a time. The other subtask, Floating (3), only appears easy because no pupils reached the level scored highest, namely noticing two features (the effect of two physical quantities, weight and density, at the same time). The 'high' score is thus not comparable with other scores of the 'same' level.

Rolling (12) and (3):

Here pupils had to notice the actual paths of a ball rolling. In subtasks (12) that the path should be visibly parabolic, and in the other subtask (3) that the path could appear straight downwards or bent - as part of a parabola. To score highly, pupils had not to 'see' any path which, from other research (McDermott, 1983), it is known that they may well wrongly expect to see. Thus scores on 'noticing' seem likely to be affected by children's expectations.

Balancing:

Here, noticing the position of the balance was easy, and to make the scores discriminating, the higher score was given only for noticing plus giving some reason.

Black Boxes:

In this case, the noticing involved discriminating longer or shorter distances inside boxes, from feeling and hearing moving the ball inside. The high score for the latter required making a rather fine discrimination - like the one required if they attempted an analytical procedure to solve the problem in I + G -, based on the information given by the drawings.

Analysis:

Overall, then, on this account it may seem that 'noticing', as defined, is rather varied in its nature.

Table 7.1.1 shows 2 x 2 tables for noticing, for each pair of tasks (or subtasks in the case of Floating and Rolling).

		B	BB	F12	F3	R12
BB	0 | 10 2 | 14	7 3 10 4				
F12	0 | 12 2 | 12	10 2 7 5	6 6 4 8			
F3	0 | 6 1 | 18	4 2 13 5	0 6 10 8	WITHIN		
R12	0 | 19 2 | 5	15 4 2 3	8 11 2 3	9 10 3 2	3 16 3 2	
R3	0 | 15 2 | 9	11 4 6 3	6 9 4 5	9 6 3 6	4 11 2 7	WITHIN
		17 7 0 2	10 14 0 2	12 12 0 2	6 18 0 1	

Code: the number 12 indicates the combined subtasks 1 & 2; 3 indicates subtask 3.

Table 7.1.1: contingency tables for 'noticing'.

Floating 3 presents the special problem that most pupils scored 1, which, as explained above, represents only a modest performance. It has little power to discriminate - although it enables one to see the difficulty of noticing two physical quantities at the same time - and indeed in every case the relationship with the other tasks is essentially zero or has a negative sign. In each case, the largest number of pupils are those who gain 1 on F3, but gain the lower score on the other task, consistent with the account given above.

It was argued before that BB and R involve special aspects of 'noticing', in one case an auditory discrimination and in the other observing a behaviour which is not that which is expected. The data for BB are consistent with this kind of interpretation, in that in four of the five comparisons - the other one explained by its relation to F3 -, there is no evidence of any relationship with noticing in the other tasks. A similar result appears for R12 and 3, but here the lack of relationship might be attributed to the special effects of the interference of expectation.

It is clear that no great weight can be attached to such arguments purporting to account for a lack of relationship, which might just as well be attributed to random variation. The reason for drawing attention to possible explanations, in terms of features of the tasks, is to highlight the difficulty in practice, in realistic classroom tasks, of realizing a process such as 'noticing' in any uniform way, without artificiality.

B and F12 are, on the face of it, more comparable tasks in respect of 'noticing'. The sign of the relationship is indeed positive - though the value of $\ln \alpha$ is not quite twice its standard deviation. In both tasks the situations imply noticing the effects of different physical quantities: weight and distance in the case of Balancing, and weight and density in the other.

7.1.1.2 Making generalizations.

On the face of it, 'making generalizations' seen as a process appears to have a good deal in common in all the tasks. The nature of the answers is restricted to making none, one or two plausible generalizations. The collapsing of the scores could also be done in a way that maintained comparability, and at the same time so as to make the marginal totals not too different.

An example of the kind of generalizations that were counted as plausible in the Balancing task can be taken from Chapter 8. Three types of generalizations were made. They were: 1) the more weight there is on one side, the more it tilts (given that distances are equal); 2) If one weight is closer in, one which is further out needs to be less, so as to compensate; 3) The balance must always have the same amount of bricks or weight on either side. Though they vary in generality, these are all generalizations, and appear comparable with the generalizations found in the other tasks (Chapter 8).

Table 7.1.2 shows that all the relationships have a positive sign. None are negative, and the pattern is consistent. In all cases, fewer than expected children do well in Floating and less well in the other (easier) task. This may be connected with the difficulty of seeing density as a variable, preventing children from making a plausible generalization.

			BT		BBT		FT	
BBT	0	9	5	4				
	1	15	7	8				
FT	0	15	9	6	7	8		
	1	9	3	6	2	7		
RT	0	9	5	4	5	4	8	1
	1	15	7	8	4	11	7	8
			12	12	9	15	15	9
			0	1	0	1	0	1

Code: T indicates that all subtasks were combined

Table 7.1.2: contingency tables for 'making generalizations'.

The difficulty of using significance tests with these data is illustrated by the relationship between Rolling and Floating, which is the strongest one. With a = 9.1 and $\ln \alpha = 2.2$ the relationship looks convincing, but at the same time the estimate of the standard deviation of $\ln \alpha$ is 1.2, which is still relatively large. However, the fact that all relationships are positive lends some support to the hypothesis that 'making generalizations' is a broadly similar process in all the tasks.

7.1.1.3 Identifying variables.

There were two difficulties in constructing process scores for the tasks, in this case. In Black Boxes, by the nature of the task, the best that could be done was to count as variables such things as 'the pattern inside the box', or 'roll the ball along the left and right hand sides'.

In Roller Ball, the quantities which had to be counted as variables are more like manipulations (pulling the spring or pointing the gun). In this case, in addition, nearly all children were given the lower score, because they failed to give justifications - relying only on the manipulation built into the task. Thus this score cannot discriminate between children.

In Balancing and Floating, genuine variables were available to be identified and performances were such that discriminating scores could be constructed.

The pattern of relationships in the contingency tables (Table 7.1.3) tends to agree with this account. There is a positive relationship between Balancing and Floating (ln $\alpha = 2.7$, standard deviation 1.2). The relationships between Black Boxes and Floating or Balancing are weak or zero, corresponding to the artificial nature of the 'variables' defined for Black Boxes. None of the relationships of Rolling with the other tasks can be relied upon, due to the unbalanced nature of the marginal totals.

			BT	BB1	FT
BB1	0	12	8 4		
	1	12	8 4		
FT	0	12	11 1	7 5	
	2	12	5 7	5 7	
RT	0	21	15 6	11 10	11 10
	2	3	1 2	1 2	1 2
			16 8	12 12	12 12
			0 2	0 1	0 2

Code: T indicates that all subtasks were combined; BB1 = Black Boxes subtask 1, with subtasks 2 & 3 not considered (see Chapter 5).

Table 7.1.3: contingency tables for 'identifying variables'.

In summary, an argument can be made for the process 'identifying variables' as comparable between tasks in the case of Balancing and Floating, where variables are most clearly present in the nature of the tasks. Although the comparability between tasks is essential for the notion of process, one should not forget the nature of the answers given in this context. Some of them, like 'the colour of the water' [for density in the Floating task] or 'the amount of bricks' (for weight in Balancing) are reminders of the differences between what some might count as identification of variables, and what children actually do.

7.1.1.4 Imagining causes.

Differences in nature are clearly present in this 'process' as between tasks. Two tasks, Black Boxes and Balancing , are different in nature but for quite different reasons. Floating (13) and Rolling look similar. The other one, Floating (2) leaves practically no room for relationships given its very low power of discrimination. Black Boxes and Rolling also show rather unbalanced marginal totals - the first quite easy, the second quite difficult -: these leave more space for relationships than Floating (2), but still have little power of discrimination.

Floating (2):

Results show how difficult it is to imagine agents of causation, responsible for floating and sinking, other than heaviness and lightness. They also show the problems of comparability between processes due to the fact that there are no children performing at what was considered the highest level.

Black Boxes:

Imagining a cause for this task is restricted to 'identifying' a cause or not in one question, while Balancing and Rolling have the usual three levels of performance and two questions. The nature of the causal agent is also rather peculiar in the sense that is part of the device - 'patterns inside of the boxes' - and its effects become apparent when some manipulations are made - to move the box in a particular way.

Balancing:

This task seems to have a 'truly' imaginative element in its nature. It is the only task where the causal explanation - fingers and bricks both 'push down' - cannot be seen; with senses having little role to play.

Floating (13) and Rolling:

In both cases the nature of the causal agents seems to be similar. Although the causal element is 'built in' the tasks - water is in the beaker and 'pushes up'; the board is tilted and 'produces' parabolic trajectories -, it does not depend on direct manipulations to show its effects and does not require a substantial imaginative effort. It is not self evident that 'inert' elements such as the ones described, can be responsible for phenomena. Floating (13), had its scores based on two questions, both dichotomous, allowing responses like 'amount of water' to be counted as 'high', just because it gave water a role to play.

Analysis:

In table 7.1.4 Black Boxes has just one strong relationship, with one empty cell - no child does badly on Black Boxes and does well on Rolling. The cause in both tasks is 'built' in - something that might explain their relationship -, but the 'behaviour' of Black Boxes is not consistent with Floating - where the cause is also 'built' in. Thus, no obvious interpretation seems possible for this result. Balancing, according to the explanations given, has an 'odd' behaviour; showing only negative relationships with all the tasks, perhaps due to its somehow imaginative nature - fingers and bricks 'pushing down' -, rather than focusing their attention into the device itself - gravity is not 'built' in the task. Floating (13) and Rolling, except for their negative relationships with Balancing and their different 'behaviour' with Black Boxes, have a weak relationship (ln $\alpha = 0.8$ and standard deviation 0.9).

			BT 0	BT 1	BB 0	BB 1	F13 0	F13 1	F2 0	F2 1
BB	0	7	3	4						
	1	17	14	3						
F13	0	10	5	5	3	7				
	1	14	12	2	4	10				
F2	0	22	15	7	7	15	WITHIN			
	1	2	2	0	0	2				
RT	0	17	11	6	7	10	8	9	17	0
	1	7	6	1	0	7	2	5	5	2
			17	7	7	17	10	14	22	2

Code: T indicates that all subtasks were combined; F13 = Floating, subtasks 1 & 3 combined; F2 = Floating, subtask 2; BB has just one subtask.

Table 7.1.4: contingency tables for 'imagining causes'.

Given the diversity in nature -but not artificiality - in some cases and the lack of comparability in others, this 'process' seems to be reduced to the similarity shown by Floating (13) and Rolling; where a 'hidden' causal element is responsible for phenomena like floating and parabolic trajectories. Doubts might be raised in relation to the contradictory 'behaviour' of Black Boxes. As a whole if any such process exists it seems to be dependent on the nature of the causal agent.

The analysis of previous paragraphs have been based on the nature of what has been elicited and the sign of the relationships between the same process (in this case 'imagining causes') in different contexts, but what about the statistical grounds in which most studies of processes are rooted. These statistical bases are in part 'responsible' for the existence of science processes. As Ogborn (1990) says, there are two conditions that need to be satisfied for us to recognize a 'process' as existing in people's behaviour: "that we have criteria to recognize it which can be stated in a relatively context independent form, and that it turns out that we can recognize this process in different contexts". But, the possibility of such recognition implies variability: "we must know what it is for people not to be able to use a certain process, to be able to distinguish cases when they can" (Ogborn, 1990). The correlational approach, that of looking to see whether a process exist across different contexts, is based on the same idea. But, as Ogborn (1990) says, "a correlation uses only variation *within* the sample studied, in this sense correlational studies are norm-referenced". This leads in his view to a seeming paradox: "if we study a process or skill possessed by all or most of the sample, we will be unable to say purely from correlation data whether it is used across contexts". Thus, "everything now depends on whether the criteria look sound: that is, whether we are satisfied with a criterion-referenced formulation. To have this confidence we need at least a commonsense notion of what it would be for children not to use the process or possess the skill. If - and only if - we have that confidence, then data which suggests that all or most children have a given competence and that is repeated in different contexts, can be interpreted as meaning that the competence exists in different contexts, despite the impossibility of producing correlational data" (Ogborn, 1990).

There are signs that the general case described by Ogborn for correlational studies, may well have particular applications. In Chapter 6 (Figure 6.9) 'imagining causes' was described as as the most difficult 'process' (from those studied here) if only the number of children performing 'not good' (60% of them) is taken into account; reducing considerably the variability of the data. In this chapter (Table 7.1.4) it can be noticed that there are three cases (B, F2 and R) out of five, where dichotomization is very uneven in favour of value Ø; making it very difficult for a strong correlation to appear. This results seems to suggest that in strict correlational terms we would not be allowed to give statistical existence to 'imagining causes'. But, it leaves the door open to investigate whether in a longitudinal study the process 'imagining causes' (or others) tend to show more or less variability according to children's age, and the logical criteria for its existence; as Johnson (1987) says 'the reliability coefficients quoted in the literature are appropriate for norm-referenced assessment, where maximum discrimination among pupils is the aim. For diagnostic tests and criterion-related assessment these traditional measures are inappropriate'; similar point of view is hold by Lang (1982).

7.1.1.5 Understanding.

Here comparability between tasks again presents a problem, due to the varied nature of the 'processes' as defined. To overcome this difficulty, a decision concerning scoring and conceptualizing the categories was made. It was to collapse scores in such a way that zero was given for the lowest level of understanding in each task, leaving medium levels of performance (including rather unsatisfactory understanding of what was involved) and high scores under one collapsed category. This was done because although the nature of 'understanding' seemed rather variable across the tasks, it allowed taking the lowest scores in each case as a level which arguably consistently represents a lack of understanding.

Balancing:

Children with the lowest score think of weight always in discrete terms, with the rest thinking of it at least sometimes in continuous terms.

Black Boxes:

In this task what discriminates children's performance is the strategy they think is the best for finding out which box is which. The score is low when children select a box and do not abandon it until they can match it to one of the drawings; medium and high scores are for those who think that selecting a picture and matching it with one of the boxes or selecting a test to compare boxes are the best options. It is of course, debatable whether this difference should count as 'understanding'.

Floating:

The factors responsible for different levels of floating in the task, are what discriminate children's answers. The lowest is when among the factors selected is amount of water, with the rest selecting either weight or kind of water or both.

Rolling:

The plausibility of different trajectories is what divides children's performance into different categories. The low score in this case is when children select only non-plausible paths, with higher scores for selecting either a mixture of plausible and not plausible, or selecting only plausible ones.

Analysis:

In Table 7.1.5 it can be seen that there does not seem to be a consistent pattern of relationships. The Balancing task, for which it can be argued that the process is artificial, has in fact two of the three positive relationships. Although Floating and Rolling have a positive relationship, Balancing and Rolling have a negative one and Balancing and Floating have very little relationship.

It is interesting in the light of the above reservations, to note that the shape of the relationship of Floating with all the other tasks is as if success in it were a condition for good performance in the others. All three relationships have very few cases where children did badly in Floating and well in the rest; suggesting that children who believe that the amount of water is responsible for different floating levels, cannot overcome more difficult conceptualizations like thinking of weight in continuous terms, to start with giving some internal structure to a box or imagining plausible paths.

As a whole it seems that, if such process as understanding exists, it is dependent on the nature of what is to be understood; it suggests that children develop their own frameworks, as Driver (1983) does.

			B		BB		F	
BB	0	12	8	4				
	1	12	8	4				
F	0	12	11	1	7	5		
	2	12	5	7	5	7		
R	0	21	15	6	11	10	11	10
	2	3	1	2	1	2	1	2
			16	8	12	12	12	12
			0	2	0	1	0	2

Table 7.1.5: contingency tables for 'understanding'.

7.1.1.6 'What-if' reasoning.

Comparability, from the point of view of the collapsing system, was not a problem. Some tasks vary in nature, others look similar. Rolling, Black Boxes (p) and Black Boxes (t) look different in nature and, Balancing, Floating and Black Boxes (hr) look similar.

Rolling:

This was intended as a prediction of what would happen if the ball is rolled from a certain position towards a trap, and it turned out to be rather influenced by expectations; making high levels of performance quite difficult, and discrimination also difficult. In nature it does not match the reasoning found for Floating, Balancing and Black Boxes (hr), because it deals with predictions of trajectories rather than values of variables (velocity or direction). This because it was found difficult for children to read angles, so the idea of a functional relation between direction and velocity (knowing if the ball goes into the trap or not) was abandoned, opting for a notional prediction (trajectory) instead.

Black Boxes (p):

Here children had to set the ball in all four corners of a box knowing the features inside the boxes. Although it requires reasoning of a 'what if' kind (based on the information given by the drawings), it again seems to be different in nature from the others, because of being dependent on practical manipulation.

Black Boxes (t):

Children are required to make use of 'what-if reasoning' to devise a test to find out which box is which. This, however, is an application of what-if reasoning to devise a test, and not the use of what-if reasoning itself.

Balancing, Floating and Black Boxes (hr):

These tasks seem comparable from the point of view of the collapsing system and from their nature. In all of them children make none, one or two what-if arguments of the kind described in Chapter 8, with scores 0, 1 and 2 respectively. In the case of Balancing what-if reasoning consists of predicting the weight or distance required to get the balance level, given certain conditions. Floating requires predicting different floating levels given certain restrictions. Black Boxes (hr) deals with telling in advance which box(es) it is possible or not to know about, given certain conditions. All these tasks have fairly equal powers of discrimination, as can be seen from Table 7.1.6.

			BT	BBp	BBThr	BBt	FT
BBp	0	12	9 3				
	2	12	6 6				
BBThr	0	11	9 2	WITHIN			
	2	13	6 7				
BBt	0	14	9 5	WITHIN	WITHIN		
	2	10	6 4				
FT	0	10	9 1	7 3	7 3	7 3	
	2	14	6 8	5 9	4 10	7 7	
RT	0	22	14 8	12 10	11 11	12 10	9 13
	2	2	1 1	0 2	0 2	2 0	1 1
			15 9	12 12	11 13	14 10	10 14
			0 2	0 2	0 2	0 2	0 2

Code: T indicates that all subtasks were combined; BBp = Black Boxes, setting marbles practically; BBhr = Black Boxes, hypothetical reasoning; BBt = Black Boxes, devising a test.

Table 7.1.6: contingency tables for 'what-if reasoning'.

Analysis:

In this Table it can be seen that Balancing, Floating and Black Boxes (hr) all show signs of quite strong positive relationships. The strongest is between Floating and Balancing where $\ln \alpha = 2.48$ and $\delta = 1.2$. The other quite strong ones, Black Boxes (hr) with Balancing and Black Boxes (hr) with Floating, have $\ln \alpha = 1.65$ and $\delta = 0.9$ and $\ln \alpha = 1.76$ and $\delta = 0.9$, respectively.

Results, on this line of reasoning, seem to suggest that it might be appropriate to use the same 'label' for processes belonging to these three tasks. It is difficult to think of Rolling as part of 'what-if reasoning' because of its lack of discrimination and its rather special nature. It is also difficult for Black Boxes (p) and Black Boxes (t) to be thought of as part of the same process due to their different nature and to the weakness of the relationships shown with the other tasks.

7.1.1.7 Processes across tasks.

Overall, the results of 7.1.1.1 to 7.1.1.6 suggest that it is possible, but often difficult, to devise tasks and questions within those tasks, to elicit similar processes between tasks. The two tasks which most consistently showed some similarity of processes were Floating and Balancing. The Rolling task, different in that success depended much more on manipulative skill, seems to offer less possibility of eliciting processes common to other tasks. Equally, the 'content free' Black Boxes task appears often to behave differently from the others.

As for processes, 'making generalizations' is the one for which the most convincing case can be made, or, to put it another way, is the one for which it turned out that comparable questions eliciting comparable answers could be devised. Similarly, 'what-if reasoning', which appears on the face of it to be rather like generalizing, providing some comparability between tasks.

By contrast, 'noticing' appears to be very largely task dependent, and there is some reason to suppose that the differences are conceptual in origin (as opposed to perceptual). The same position is adopted by the APU (1988): 'performance outcomes should be interpreted as the product of an interaction between process and content (where 'content' implies theory-laden assumptions ...'). It is notable that 'understanding' was very variable across tasks, though it can also be criticised for being ill-defined. 'Identifying variables' though similar in Floating and Balancing, also seems to depend on children's expectations.

This whole picture seems to indicate,

i) that it should not be taken as unproblematic to define a process which is the 'same' in different tasks,

ii) that the content of the tasks, and children's understanding of it is very important,

iii) that where 'processes' do seem to be similar, they are conceptual rather than perceptual.

7.1.2 WHAT WOULD BE EXPECTED FROM PROCESSES WITHIN A TASK?

It has been previously discussed whether it makes sense to give the name processes, to some behaviours of children; depending on the nature of what is being elicited and their consistency in nature and performance across tasks. It was found that there are some grounds for labelling some performances as processes, while others are better described as dependent on the content of the task.

Here the discussion will be centred on to what extent processes are related within a task depending on the nature and consistency shown by such processes in the previous analysis (see section 7.1.1).

7.1.2.1 Balancing.

There seem to be reasons to suppose that noticing has a cognitive connotation. In addition performance, as defined, was discriminated in terms of whether or not a reason was given for what they observed. Therefore relationships might be expected with performances that also seem to be essentially cognitive, like 'what-if reasoning', 'identifying variables' and 'making generalizations'. Such cognitive processes might be expected to relate positively with one another. In the case of 'imagining causes' it is expected that not many relationships would be found because of its 'truly' imaginative nature (see section 7.1.1.4). The relationships found are in Table 7.1.7 .

	WT 0	WT 1	VT 0	VT 1	GT 0	GT 1	
NT 0	5	6	2	9	7	4	11
NT 1	6	7	0	13	5	8	13
	11	13	2	22	12	12	

	WT 0	WT 1	VT 0	VT 1	
GT 0	8	4	2	10	12
GT 1	3	9	0	12	12
	11	13	2	22	

	WT 0	WT 1	
VT 0	2	9	11
VT 1	0	13	13
	2	22	

	NT 0	NT 1	U 0	U 1	WT 0	WT 1	VT 0	VT 1	GT 0	GT 1	
CT 0	6	11	9	8	9	8	2	15	7	10	17
CT 1	5	2	6	1	2	5	0	7	5	2	7
	11	13	15	9	11	13	2	22	12	12	

Code: T indicates that all subtasks were combined; W = What-if reasoning; V = Identifying variables; G = Making generalizations; N = Noticing; C = Imagining causes; U = Understanding.

Table 7.1.7: relationships between 'processes' in Balancing.

It can be seen from the Table 7.1.7 that 'noticing', in agreement with this argument, has a positive relationship with 'making generalizations' and 'identifying variables', although the latter is weak. No relationship was found with 'what-if reasoning'. This may be not surprising taking into account that the reasons given in noticing mostly refer to situations where evidently there are more weights on one side and the balance is tilted, and that 'what-if' reasoning demands understanding of the possibility of getting the balance level having more weight on one side, given a distance which is not the same.

'Making generalizations' relates quite strongly with 'what-if' reasoning ($\ln \alpha = 1.8$ and $\delta = 0.9$) and, weakly but positively with 'identifying variables'. The relationship between 'what-if' reasoning and 'identifying variables is also weak and positive. All are processes which were argued above to be cognitive in nature.

Given the imaginative nature of 'identifying' the cause in this task - gravity as an element not 'built' in the device or visible - not many relationships were expected. What was found was a pattern of negative relationships, except for a pair of weak positive ones with 'what-if' reasoning and 'identifying variables', that seems to confirm its 'odd' nature.

7.1.2.2 Floating.

For the same reasons as with Balancing, 'noticing' might be expected in the Floating task to have positive relationships with identifying variables, what-if reasoning, making generalizations and understanding, which are also expected to be related to one another.

	VT	C13				
U	8 4 4 8	8 4 8 4	12 0 12 2			
	12 12 0 2	16 8 0 2				
	WT	VT			WT	
GT	7 8 0 9	4 11 0 9	15 0 9 1	VT	9 3 1 11	12 0 12 2
	7 17 0 1	4 20 0 1			10 14 0 2	
	VT	WT	GT		U	
N12	6 6 6 6	6 6 4 8	12 0 11 1		7 5 5 7	12 0 12 2
	12 12 0 2	10 14 0 2	23 1 0 2		12 12 0 2	

Code: T indicates that all subtasks were combined; V = Identifying variables; C13 = Imagining causes, subtasks 1 & 3 combined; U = Understanding; W = What-if reasoning; G = Making generalizations; N12 = Noticing, subtasks 1 & 2 combined.

Table 7.1.8: relationships between 'processes' in Floating.

Table 7.1.8 shows that noticing has a positive relationship with all the processes considered as 'cognitive', but that all are weak. Making generalizations seems to have a strong relationship with what-if reasoning: no children who did badly on what-if reasoning did well in making generalizations. And also making generalizations seems to have a strong relationship with identifying variables: no child who did badly in identifying variables, did well in making generalizations. To complete the relationships between the cognitive processes, identifying variables and what-if reasoning show signs of a strong relationship (ln $\alpha = 3.49$ and $\delta = 1.23$); as if making generalizations depends on both identifying variables and what-if reasoning.

7.1.2.3 Rolling.

Noticing is again expected to relate to understanding - what kinds of paths are plausible or not - and to what-if reasoning - prediction of the path. Identifying variables and making generalizations, as cognitive processes, might be related.

	VT 0	VT 1	Total	GT
GT 0	4	5	9	0
GT 1	7	8	15	1
Total	11	13		
VT	0	1		

	UT 0	UT 1	WT 0	WT 1	Total	N12
N12 0	10	2	11	1	12	0
N12 1	5	7	8	4	12	1
Total	15	9	19	5		
	0	1	0	1		

Code: T indicates that all subtasks were combined;
V = Identifying variables; G = Making generalizations
U = Understanding; W = What-if reasoning; N12 = Noticing, subtasks 1 & 2 combined.

Table 7.1.9: relationships between 'processes' in Rolling.

In this case it was argued previously that noticing is strongly affected by pupils' expectations. Table 7.1.9 shows that this idea has some support from the strong relationship between noticing and understanding (ln $\alpha = 1.94$ and $\delta = 0.97$) and a mild positive relationship with what-if reasoning (predicting). The two cognitive processes, making generalizations and identifying variables, seem however not to be related to one another.

7.1.2.4 Black Boxes.

Noticing, by its nature in this task, does not have a cognitive connotation, but may be related to identifying variables and imagining causes; comparing distances between walls might give an idea of what to pay attention to and of what could cause different paths for the ball when moved inside of the boxes. The strategy -Understanding - that suits better how to find out which box is which, might relate to devising a test (Dt), as both have the same purpose: to find out which box is which. What-if reasoning (HRT for hypothetical reasoning total) would relate to what has been considered as a cognitive process, making generalizations; as well as to setting the ball in all four corners (P), devising a test (Dt), understanding (U) and identifying variables (V).

U \ Dt	0	1	Total
0	6	9	15
1	2	7	9
Total	8	16	

NT \ V	0	1
0	2	0
1	10	12
Total	12	12

NT \ C	0	1	Total
0	1	1	2
1	6	16	22
Total	7	17	

HRT \ GT	0	1
0	3	2
1	6	13
Total	9	15

HRT \ U	0	1
0	5	0
1	10	9
Total	15	9

HRT \ V	0	1	Total
0	4	1	5
1	8	11	19
Total	12	12	

HRT \ P	0	2
0	8	3
2	4	9
Total	12	12

HRT \ Dt	0	2	Total
0	8	3	11
2	6	7	13
Total	14	10	

Code: T indicates that all subtasks were combined; Dt = Devising a test; U = Understanding V = Identifying variables; C = Imagining causes; N = Noticing; G = Making generalizations; P = Setting marbles practically; HR = Hypothetical reasoning.

Table 7.1.10: relationships between 'processes' in Black Boxes.

Table 7.1.10 shows the actual relationships. The expected relationships for noticing are generally positive, although weak. Devising a test (Dt) and following a good strategy seem to have a mild relationship. What-if reasoning in this task (HRT) relates positively with all the processes mentioned. Only in the case of its relationship with setting the ball in all four corners (P) does there seem to be a strong relationship (ln $\alpha = 1.79$ and $\delta = 0.90$), although it also shows a mild relationship with devising a test (Dt). It is also interesting to note that nobody who did badly in HRT did well in selecting a best strategy other than trying to match a box with one of the drawings. The 'behaviour' of identifying variables (V) is of interest if one compares the very weak relationship with noticing (NT) and its stronger relationship with HRT; suggesting that what to pay attention to or identifying variables might have a cognitive connotation, despite its odd nature (see Section 7.1.1.3).

7.1.2.5 Processes within a task.

Processes seem to have different weights within each task; that is some process(es) play a key role in one task but not in others.

The strongest relationship in Balancing is between making generalizations and what-if reasoning. This is understandable in that both require grasping how distance and weight are related.

In the case of Black Boxes, the process that seems to dominate the task is what-if reasoning (HRT). This is particularly encouraging, the task being one which is specifically designed to require hypothetical thinking.

In Floating, as in Balancing, are the cognitive processes which seem to play a key role. Being able to identify the factors affecting floating and sinking, knowing the effect of one upon the other and making generalizations are all related. Again, this makes sense in view of the nature of the task.

The final task, Rolling, reflects its difference in nature from the other by having noticing, understanding and predicting (what-if reasoning for this task) as the main processes. Noticing relates strongly to understanding and also, but less so, to predicting. These relationships seems to show the role children's expectations play in this task; at least in the way it was devised.

As a whole it can said that,

i) Processes have different weights in each task, though with Floating and Balancing having similar processes playing a key role.

ii) Cognitive processes like making generalizations, what-if reasoning and identifying variables, tend to be the most important.

iii) Processes like noticing, understanding, and imagining causes, seem not to have a key role in these tasks. The exception is Rolling, where children's expectations seem to give noticing a more important role.

iv) A task specifically designed to elicit 'what-if' reasoning does in fact appear to do so.

7.2 INVESTIGATION PLUS GOAL VERSION OF THE TASKS.

7.2.1 WHAT IS THE NATURE AND EXTENT OF PROCESSES?

Comparisons of processes across tasks will be looked at in the same way as the processes belonging to the structured versions of the tasks: qualitatively and via 2 x 2 contingency tables.

7.2.1.1 Searching.

The nature of the search space for each task and the differences in dichotomizing the scores will be explained. This 'process' indicates the extent to which the child tries all possibilities in a task, seen as a search in a space of possibilities (see Chapter 5).

Balancing:

Searching through the space of this task means to try possible combinations of 'values' of two variables, in this case weight and distance. They are values in the sense that they reflect different types of 'bricks' - identical, equivalent, ... - and 'stripes' - equal distance, ... -, but not in a numerical sense (see Chapter 5).

Floating:

In the way this search was defined (see Chapter 5), again nearly all children scored highly, and no relationship can appear.

Black Boxes:

Two ways of scoring were tried (BB1 and BB2). Neither works well, either recording all children as high scoring, or all as low scoring. This is because all children tried all the boxes (high BB1) but none compared boxes (low BB2). Thus these scores cannot give any useful relationships here.

Rolling:

The space here consists of the relative positions of 'gun' and 'trap'. Scores here, as in Balancing, are fairly evenly divided.

Table 7.2.1 shows the contingency tables. The only case where a relationship can be considered is for Balancing and Rolling. The relationship is a positive one but not strong.

			B	BB1	BB2	F
BB1	0	--	--- ---			
	1	24	11 13			
BB2	0	24	11 13	--- 24		
	1	--	--- ---	--- ---		
F	0	1	1 0	--- 1	1 ---	
	1	23	10 13	--- 23	23 ---	
R	0	11	6 5	--- 11	11 ---	1 10
	1	13	5 8	--- 13	13 ---	0 13
			11 13	-- 24	24 --	1 23
			0 1	0 1	0 1	0 1

Code: BB1 = all boxes tried; BB2 = comparison between boxes.
Table 7.2.1: contingency tables for 'searching'.

Thus we may say that the concept of 'completeness of search' while able to let us see the manner in which children in fact understood the tasks, was not in general able to be implemented in such a way as to let us see relationships between tasks.

7.2.1.2 Controlling variables.

Scores for this 'process' try to indicate to what extent children manipulate only one variable at a time.

In these cases, Balancing, Floating and Rolling, all but one or two children obtained the same score on a given task. Thus Table 7.2.2 cannot contain any relationships. We can see how children performed on each task, but not how their performances on different tasks are related.

			B	F
F	0	1	0 1	
	2	23	1 22	
R	0	3	0 3	0 3
	2	21	1 20	1 20
			1 23	1 23
			0 2	0 2

Table 7.2.2: contingency tables for 'controlling variables'.

7.2.1.3 Replicating.

Scores for this 'process' reflect the degree to which children made attempts to repeat a phenomenon under the same conditions in order to check what happened.

Balancing and Floating:

In both cases, almost no child did any replicating; neither checking the balance or the weights nor checking the floating level.

Black Boxes:

Here there was some variation in replication, which here consists of repeating a test on the same box.

Rolling:

There was variation in replicating, with children most often repeating shots with the same direction.

The two cases where there was variation in replication are both ones where the nature of what was replicated seems to be very much a feature of the task as set (manipulation of a hidden ball, or trying to fire a spring loaded gun consistently) (Table 7.2.3). As it happens, the relation between the two is negative.

			B		BB		F	
BB	0	15	15	---				
	1	9	9	---				
F	0	23	23	---	15	8		
	1	1	1	---	0	1		
R	0	14	14	---	7	7	14	0
	1	10	10	---	8	2	9	1
			24	---	15	9	23	1
			0	1	0	1	0	1

Table 7.2.3: contingency tables for 'replicating'.

7.2.1.4 Making notes.

Performance is classified as not making notes at all, making some irrelevant notes and making some relevant notes.

Balancing:

Children hardly show any tendency to make notes in this task.

Floating:

Notes are mostly related to writing down how many shots are needed to get the required floating level in each beaker. There is some variation in performance.

Black Boxes:

Here, children tend to make more notes. They usually draw the patterns inside of the boxes or write down the colour of their lids in correspondence to their findings; probably to make sure that they do not forget which is which.

Rolling:

A few children make notes about what happened - often with drawings - about the different positions they had made attempts from.

The marginal totals limit the possibility of seeing relationships in Table 7.2.4, although there do seem to be some around Floating. There are essentially no children who do not make notes for Floating and do make notes for the other three tasks. The necessity to remember how many ball bearings were needed to float the straw at the required level, seems rather evident; although not that evident for the other tasks, except for Black Boxes where is sensible to expect to see children making notes in order to remember which box was which.

			B		BB		F	
BB	0	14	12	2				
	1	10	9	1				
F	0	9	9	0	8	1		
	1	15	12	3	6	9		
R	0	19	18	1	12	7	9	10
	1	5	3	2	2	3	0	5
			21	3	14	10	9	15
			0	1	0	1	0	1

Table 7.2.4: contingency tables for 'making notes'.

There thus does here seem to be some case for considering that making notes is a behaviour with some degree of commonality between different tasks.

7.2.1.5 Processes across tasks.

An overall view of Tables 7.2.1 to 7.2.4 appears to suggest that eliciting natural or spontaneous processes, while children are performing the tasks in a less directive mode, is rather difficult. The sources of this difficulty seem to be varied: the lack of discrimination shown by some tasks and the nature of others.

In relation to the first source, processes were either very difficult (e.g.. 'replication') or very easy (e.g.. 'controlling variables'). It proved difficult to construct convincing and comparable means of scoring in different tasks. Where this was to some extent possible, as in the case of Rolling and Balancing, there seems to be a tendency for a positive relationship.

Making notes was the one case where the nature of the process and its power of discrimination were more comparable, giving some evidence of a behaviour similar in different tasks.

In summary,

i) It proved difficult to elicit natural or spontaneous processes exhibiting similar levels of discrimination and consistency of nature.

ii) It appears possible to manipulate the difficulty of the tasks by enlarging or shortening the the size of the search, if it is taken into account that the two most difficult tasks in both versions - Balancing and Rolling - (see Chapter 6) are also those which have the largest space to be searched. However other factors affecting the difficulty of the tasks can not be denied (e.g.. expectations in Rolling).

7.2.2 WHAT WOULD BE EXPECTED FROM PROCESSES WITHIN A TASK?

Section 7.1.1 analyzed to what extent the label 'process', considering different aspects of children's performance, can appropriately be applied. It was found that it is difficult to call such behaviours or performances processes (in the sense of something common to more than one task), due sometimes to a lack of variation in scores and in others to differences in the nature of tasks. In fact the only reasonable candidate as a process common to more than one task was the case of 'making notes', and even this may be doubted.

Given this situation, the only meaningful and feasible relationships are those between 'making notes' and 'searching'. One might expect that in order to not forget which kind of attempts have been made, it would be useful to make some notes; something expected to happen in all the tasks.

Table 7.2.5 shows that all relationships are very weak, very liable to be due to chance (Balancing and Floating) or, non-existent (Black Boxes and Rolling).

		Balancing			Floating			Rolling			Black Boxes		
		SS 0	SS 1		SS 0	SS 1		SS 0	SS 1		SS 0	SS 1	
MN	0	11	10	21	0	9	9	9	10	19	--	14	14
	1	0	3	3	1	14	15	2	3	5	--	10	10
		11	13		1	23		11	13		--	24	

Code: SS = Searching space; MN = Making notes.

Table 7.2.5: contingency tables for 'searching' and 'making notes' in all four tasks.

These results suggest that, although some expected relationships might be interesting, they turned out to be rather difficult to obtain either because of the lack of discrimination of some 'processes' or because of their differences in nature.

They should be compared with the case of the Structured tasks, where it more often proved possible to identify processes common to more than one task.

Chapter 8.
TASK AND PROCESS: WHAT DID CHILDREN ACTUALLY DO?

Evidence from both Investigation plus Goal and Structured versions of the tasks will be used here to describe how the children performed the investigations. The Investigation plus Goal tasks give evidence about what children do spontaneously, concerning which possibilities they try, how they control variables, whether they repeat observations and to what extent they make notes. Evidence about other processes, not able to be observed in the Investigation plus Goal tasks, can be obtained from the Structured tasks. These are whether they notice phenomena, how they understand the situation, how well they predict outcomes, how far they identify relevant variables, and how far they can make generalisations and identify causes. The absence of order, version and session part effects (Chapter 6) except for the Roller Ball task (a version effect), makes this combining of evidence easier. The purpose of the discussion is to use the evidence to develop a critique of the idea of 'children using scientific processes'.

8.1 HOW COMPLETE ARE CHILDREN'S INVESTIGATIONS?

The investigation plus Goal tasks require the children to investigate a limited range of possibilities, it being up to them to make the investigation as complete as is appropriate. The fundamental structure of these tasks is to offer a 'space' of possibilities to search. In Chapter 5 'search spaces' were defined for each task. A 'search space' defines all possible combinations of values of variables. These were grouped into 'configurations', which each contain a similar pattern of combinations. Staying within a configuration, a child is making changes, but staying within a particular pattern - for example keeping equal weights and equal distances on either side of the balance. This makes it possible to describe not only how completely the child investigates all possibilities, but whether they prefer some patterns to others.

Balancing:

The patterns of behaviour which emerge are:

1) the most 'obvious' combinations are avoided. Given this:

2) simple configurations are preferred to complex ones

3) symmetrical configurations are preferred to asymmetrical ones

Figure 8.1.1 shows the numbers of attempts by all children in four groups of configurations. It also shows the numbers of children who made at least one attempt within each group. The greatest number of attempts are made in the group 'single', where the same distance is used on either side of the balance, using the same or equivalent bricks. These configurations are both simple and symmetrical. However, only a quarter of these attempts use the simplest and most 'obvious' pattern with identical bricks on either side. It is of course reasonable not to try patterns where the result is evident. The 'double' configurations (using both distances on both sides) have the same symmetry as 'single' but are more complex. It can be seen that taken together 'single' and 'double' configurations represent about two thirds of all attempts. Comparing them, the more complex 'double' configurations are tried less often.

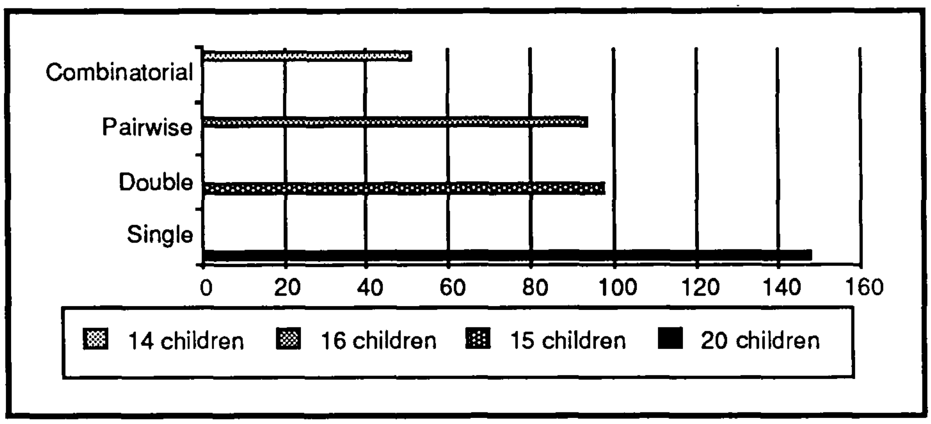


Figure 8.1.1: number of attempts made in each group of configurations indicating the number of children performing on each one of them in the Balancing task.

'Pairwise' configurations (one distance on each side, but using different distances) and 'combinatorial' configurations (combinations of more than one weight and distance on at least one side) are both asymmetrical, and are tried less often than the symmetrical ones. 'Combinatorial' configurations, more complex than 'pairwise', are tried less often.

Thus the order of frequency of attempts can be explained by a combination of preference for simplicity and symmetry. The numbers of children who make such attempts follow essentially the same pattern. A more detailed study within groups of configurations supports this interpretation. For example, in 'double' configurations, children avoid the 'obvious' pattern of exactly identical bricks on each side, but try with symmetrical patterns of equivalent sets of bricks more often than with asymmetrical patterns.

The same features can also help to account for differences in the frequencies with which children succeeded in achieving a balance. Figure 8.1.2 shows percentages of successes in each group of configurations. The greatest successes are achieved with the simplest and most symmetrical configurations. The next most frequent are those with the more complex configurations having the same symmetry, while the success rate drops to around thirty percent for the two complex and asymmetrical groups.

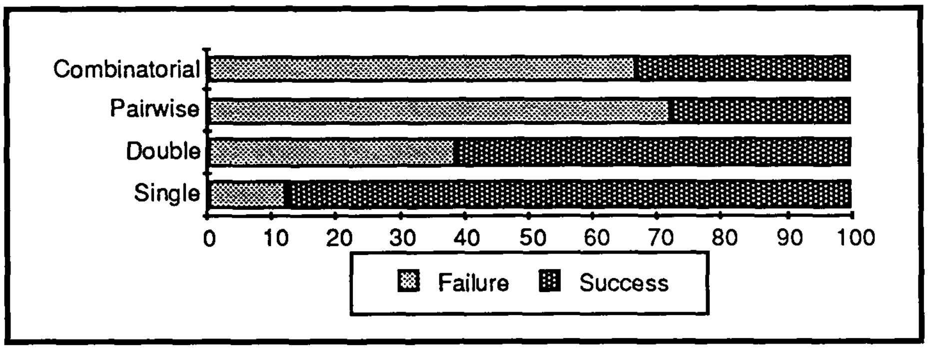


Figure 8.1.2: percentage of attempts made depending on success/failure and group of configurations in the Balancing task.

Black Boxes:

Here the notion of configurations is less useful. The main result is that children choose to try to identify the boxes one at a time, rather than comparing boxes to eliminate possibilities. All children tried all the boxes. To do that, they must have changed boxes four times. The lack of attempts to compare boxes is shown by the fact that the average number of changes of box per child is only five (see Table 3, Appendix D). It will be argued in the Conclusions that it is necessary to consider how children see the structure of situations in terms of objects or the way they isolate events, to understand how they approach an investigation.

Figure 8.1.3 shows that the patterns of trials differ between the four boxes. The differences suggest that the children were reacting sensibly to the evidence in the light of what they knew about the boxes. The total numbers of tests made on boxes C and Z are fewer than those made with boxes B and I, which can only be distinguished by a relatively difficult test (rolling the ball against an obstacle and trying to detect a gap). With boxes C and Z the three kinds of test are tried about equally often, while with boxes B and I, the particular test relevant to distinguishing them is tried more often than the others.

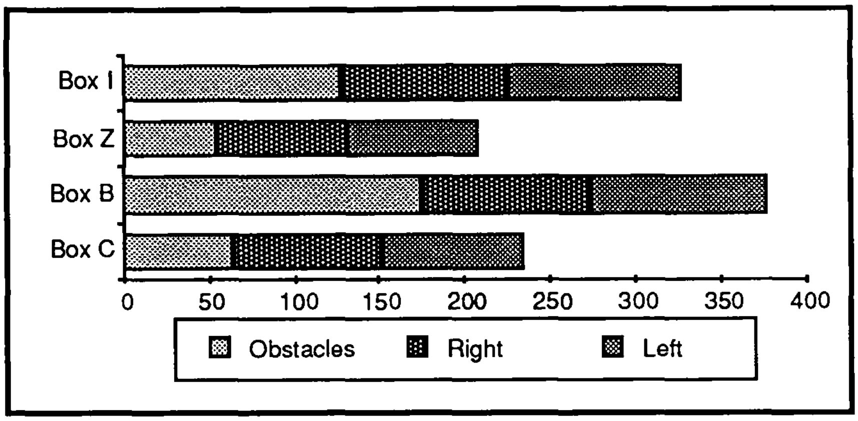


Figure 8.1.3: number of attempts made per configuration and group of configurations - box - in the Black Boxes task.

Thus it seems that here the children do have strategies, and choose them appropriately in the light of the evidence. Their overall strategy - one box at a time - is perhaps not as efficient as comparing boxes with one test at a time, but is simple and imposes little load on memory and little need to make notes.

Floating:

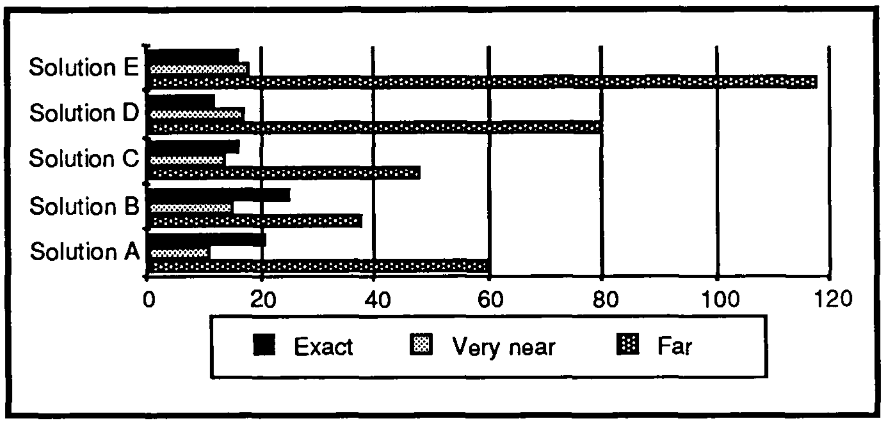


Figure 8.1.4: number of attempts made in each configuration within each solution or group of configurations in the Floating task.

In a sense, children approached the Floating task similarly to Black Boxes. All the beakers (solutions) were investigated by all children. With rare exceptions, children persisted with one beaker until they were satisfied, rather than comparing beakers using similarly loaded straws. During the investigation

47% of the attempts were spent in the same beaker changing weight and 21% changing beaker without changing weight (see Table 4, Appendix H).

In this task they had to decide for themselves what to count as a sufficient approximation to the correct number of shots to get the straw floating at the required level. Figure 8.1.4 shows that some children were satisfied with the approximation termed 'very near' (two to three shots more or less than needed), in that the number of 'exact' attempts is generally less than the number of children, and the total of 'exact' plus 'near' attempts (1.4 per child per beaker) is little more than one per child per beaker. Thus largely, children stopped when they felt they were 'near enough', without appreciable checking.

Most children tried the beakers in the sequence (A B C D E) in which they were laid out, which was in order of increasing density. That they did not know this is illustrated by their behaviour, which was generally to treat each beaker as an independent problem, and, starting with none or few shots, to add shots until the straw floated correctly. This is reflected in Figure 8.1.4 by the way the number of 'far' attempts increases from beakers B to E. The tendency in going from A to B was to use the loaded straw which had floated in A when starting B, resulting in fewer 'far' attempts in that case. This difference may be due to their not realising that it would be necessary to empty the straw to count the shots until they had tried the second beaker.

One might summarise this by suggesting that the children treated this task as five problems, not as one problem involving densities and weights. They treated the beakers as different objects, not as one entity which varied in some systematic way - rarely for example was there any compensation of weight in advance of a test. This result is discussed further below, in considering their identification of variables.

Rolling:

In this task, different configurations are different relative positions of gun and trap, and orientations of the trap. Not all configurations are equally often tried. One way of accounting for the observed differences would be to suggest that children avoid both those configurations they expect to be easy or trivial, and those which they expect to be very difficult (compare the Balance task).

As can be seen in Table 5, Appendix D, the trap is rarely tried in the easy case where it is close to the gun. If the trap is not near the gun, it appears that children see two fairly straightforward ways to get the ball into it as worth trying: firing directly, or getting the ball to fall into the trap. In Figure 8.1.5 the most common set of configurations is 'downwards' in which the ball is both fired at the trap and falls into it. The next most common is 'horizontal at bottom' which includes configurations with the ball fired directly into the trap or those in which it is fired up and falls down. Both have the trap at the bottom where a falling ball will tend to go. The least frequent configurations are those with gun and trap at the top. Further, in Table 5, Appendix D, it can be seen that configurations with the trap facing upwards (and so able to catch a falling ball) are always in the great majority, except only when the gun is below the trap. All this seems to point to children's expectations affecting which parts of the problem they try most often.

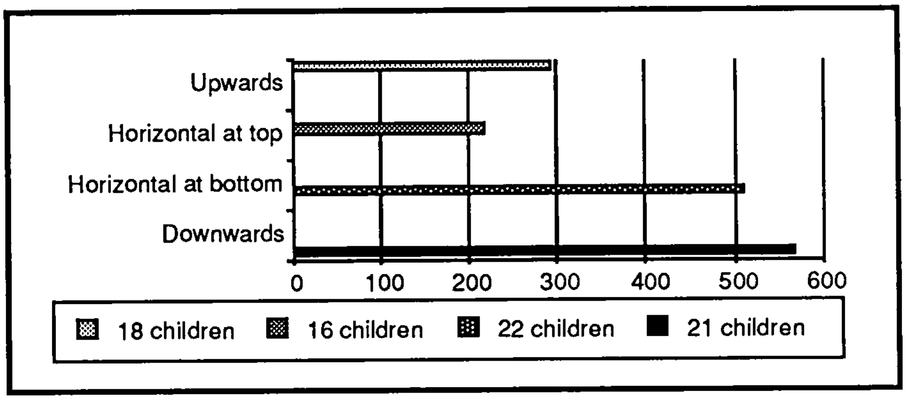


Figure 8.1.5: number of attempts made in each group of configurations indicating the number of children performing on each one in the Rolling task.

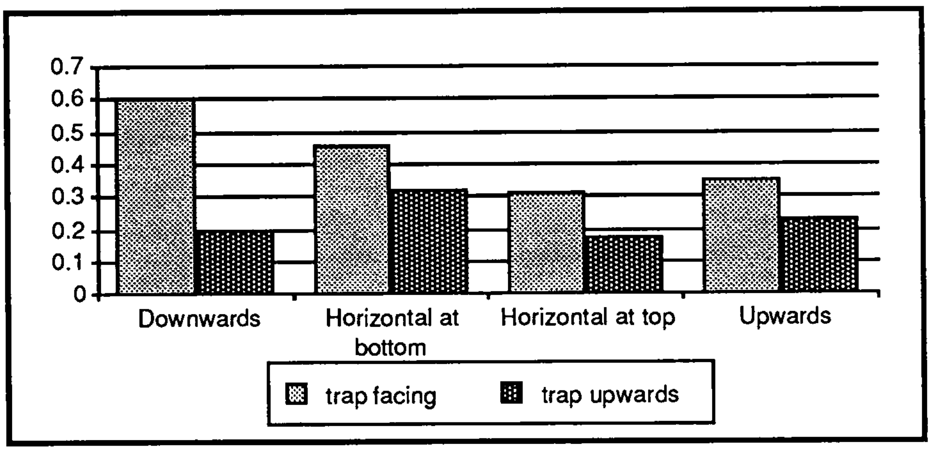


Figure 8.1.6: rate of success within each group of configurations depending whether the open section of the trap was facing the gun or not in the Rolling task.

In the rarely tried cases with the gun near the trap, the children were (not surprisingly) more often successful than with it far from the gun. More importantly, as shown in Figure 8.1.6 they generally succeeded better with the trap facing the gun than with the trap facing upwards to catch a falling ball.

Commonalities :

Overall, it appears that children treat the problem of trying possibilities differently according to whether the task itself makes it easier or harder to see what it would be to 'try everything'. In Black Boxes and Floating, the children seem to keep track of completeness using the objects in front of them, dividing the task into a number of independent tasks. In the Rolling and Balancing tasks, what has to be kept track of is not which objects have been tried but which combinations of variables have been tried. Here children try only restricted possibilities. In choosing what to do they seem to be guided by several things: by simplicity, by symmetry, and by what they expect. They seem to avoid two extremes: the 'obvious' or easy cases, and the very difficult ones.

Completeness of search :

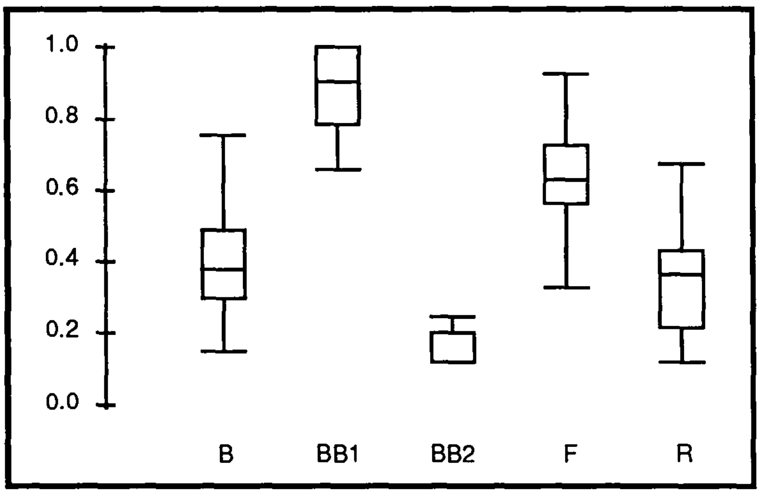


Figure 8.1.7: extent of the investigation through the search space per task.

Figure 8.1.7 shows how 'completely' children tried out all possibilities. Taking the search spaces as defined in Chapter 5, the fraction of cells in the space attempted at least once by each child was found. The boxplots in Figure 8.1.7 show how these scores are distributed for all 24 children. The differences between the tasks are due in part to differences in completeness of search but also arise in part from differences in how it was possible to define the search spaces.

The completeness of search in Roller Ball and Balancing is comparable, with a median of about 40%. Reasons for this have been discussed above. The search appears much more complete in Black Boxes and Floating, where the search can be guided by taking one object at a time. It can however be argued that these last two spaces are more artificial. In Black Boxes, a child who decides what a box is without trying all tests (which is possible) will be counted as not searching all the space. In Floating, children who are satisfied with an approximate floating level will be counted as not searching the space nearer to the exact values.

For Black Boxes, a second search space was constructed, to represent the possibilities for investigating the problem by comparing boxes using different tests in turn. The extremely limited search, looked at from this point of view, is striking.

8.2 DO CHILDREN OBSERVE OR NOTICE RELEVANT PHENOMENA AND DRAW CONCLUSIONS FROM THEM?

In Chapter 7 it was suggested that 'noticing' (as represented in this research) is rather varied in nature, and is in some cases related to expectations or to knowledge of relevant variables. Thus 'noticing' can not here be treated as anything like 'pure observation' (which few would nowadays take to exist in any case). Noticing will therefore be looked at here in relation to other aspects of the task and of the child's thinking. For these reasons, the process 'making generalisations' (Chapter 5) will also be discussed in this section. Evidence here comes from the structured versions of the tasks.

8.2.1 HOW GOOD ARE CHILDREN AT NOTICING RELEVANT PHENOMENA?

Balancing:

The two relevant tasks showed a balance tilted because an identical number of bricks were arranged at different distances on the two sides, and then level after rearranging the bricks on one side. Noticing whether the balance was level or not was not a problem, but more children gave a reason for the balance being tilted than did so for it being level (see Figure 8.2.1.1). The reasons in both cases tended to be in terms of weight rather than distance -*'I noticed that the right side was heveyer* [heavier] *then* [than] *the left side* . With the balance level, the number giving a reason is almost halved. One possible reason could be that they do not take distance into account and so find giving a reason harder when the effect of distance appears after rearranging the bricks. This fits with children's difficulty in identifying distance as a variable (see section 8.3).

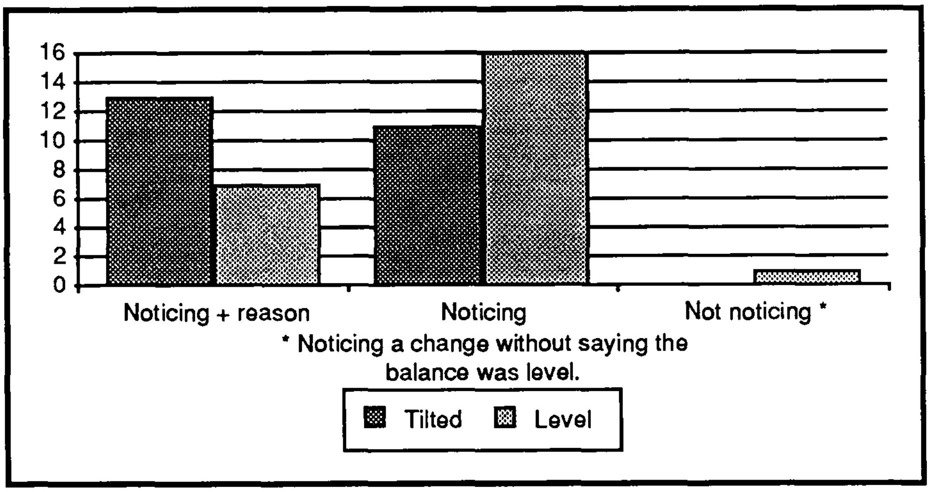


Figure 8.2.1.1: number of children noticing some phenomena in the Balancing task.

Black Boxes:

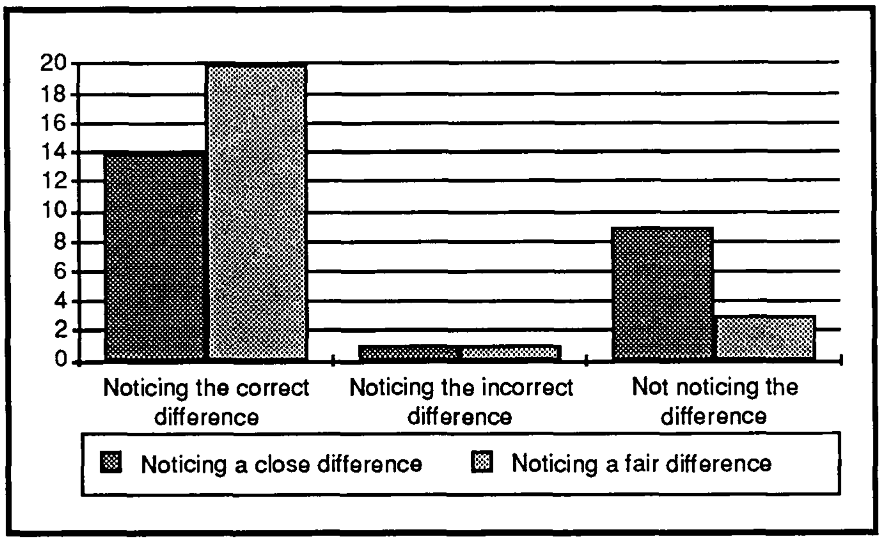


Figure 8.2.1.2: number of children noticing a ball moving across different distances in the Black Boxes task.

Here what has to be noticed has less cognitive load than in the previous case. Children have no problems in noticing substantial differences between movements of the ball in a box, but when this difference is smaller, only just more than half are able to do so (see Figure 8.2.1.2).

Floating:

Figure 8.2.1.3 shows that most had no difficulty in noticing in detail a difference in floating level when the weight was changed in water in the same beaker (same density). Only half noticed different floating levels in detail when the same weight was used in water in different beakers (density changed). Is this because they know better what to expect when the weight is changed? Certainly it seems, both in the next section and in that on causality, that the role of the water is far from clear. Their noticing changes in floating levels when changing weight coincides with their mainly identifying weight as a variable.

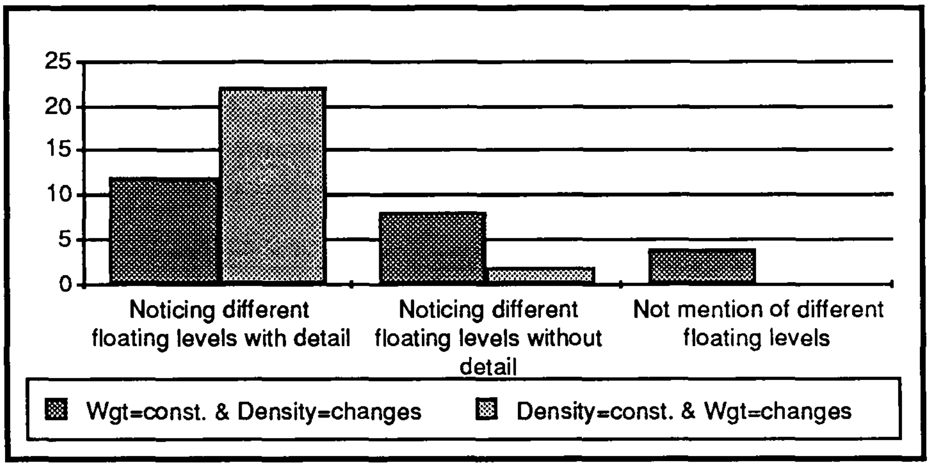


Figure 8.2.1.3: number of children noticing some phenomena in the Floating task.

When asked to look at the floating levels of two different weights each in a different density, no children noticed the effect of both factors at the same time (see Section 7.1.1.1 and Table 3 Appendix E). Most (18/24) noticed only one effect. They did notice effects of weight and density equally, however, which may be related to the use of two distinct beakers as well as two distinct weights. What had to be noticed was represented by a difference between visible objects, not just a difference between two variables.

Rolling:

In the structured Rolling task, children were asked to observe the path followed by the ball. As mentioned in Chapter 7, we know that children are likely to expect paths other than those which occur, notably a path which runs upwards and curves, but then falls straight down instead of continuing in a parabola; as the pre-Galilean theory of impetus observed by Green, McCloskey & Caramazza, 1985).

Most children 'noticed' either only implausible paths (i.e. ones which did not in fact occur) or at least some of these mixed with plausible paths (Figure 8.2.1.4).

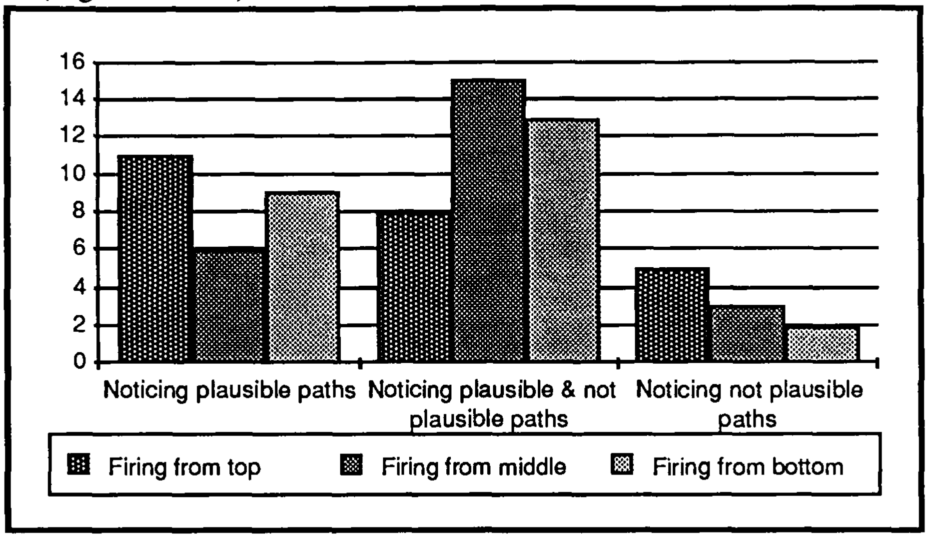


Figure 8.2.1.4: number of children noticing some trajectories in the Rolling task.

Commonalities :

Thus it seems, from the results on the Balancing, Floating and Rolling tasks, that noticing is related to 'theories', ideas or beliefs that children might have, and that it can be influenced by such expectations.

Coherence and levels of performance :

There are two cases, Balancing and Rolling 1&2, which exhibit a clear positive relationship between two questions or activities related to noticing relevant phenomena. In the case of Balancing, no children succeed in giving reasons for the balance being level having failed to give reasons for the balance being tilted (see Chapter 5). This makes sense in relation to what has been said before: that children find it more difficult to give reasons for the effect of distance rather than for the effect of weight. Another way of expressing the relationship is to say that nobody who failed in giving reasons for the effect of weight managed to give reasons for the effect of distance. In the case of Rolling, essentially no children succeed in noticing plausible paths when firing from the middle if they failed to notice plausible paths when firing from the bottom of the board (see Chapter 5). This relationship is coherent with what one might expect from such situations: it is 'easier' to notice the shape of the parabolic trajectory when firing from the bottom of the board than when firing from the middle of it - the 'gun' being pointed sideways, one can only see half of the parabola and will probably be inclined to think that on losing 'force' the ball will fall.

High levels of performance are easier when the tasks do not present, so strongly, the interference of beliefs or expectations; as in the case of Floating 1&2, Black Boxes and Rolling 3 - where 'straight' paths count as plausible - (see Figure 8.2.1.5). In the same figure it can be seen that it is more difficult to make good 'observations' when such beliefs or expectations are present; as in the case of Balancing and Rolling 1&2. Floating 3 shows the difficulty of paying attention to two different phenomena at the same time (Figure 8.2.1.5).

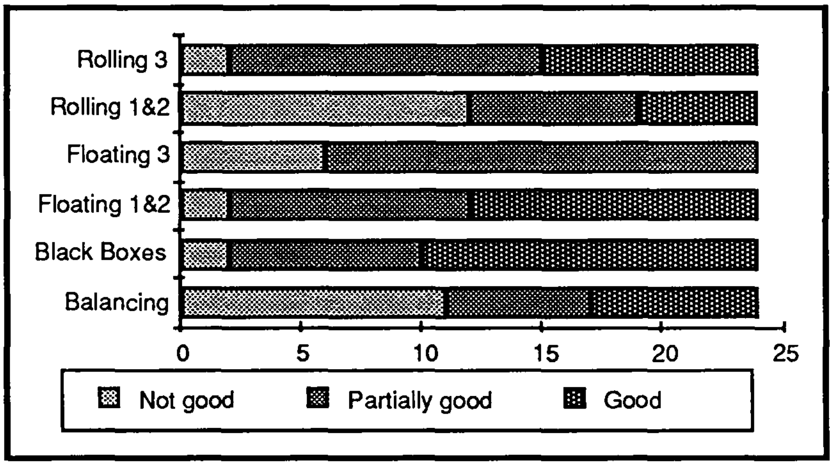


Figure 8.2.1.5: number of children scoring on Noticing Some Phenomena depending on level of performance, question and task.

8.2.2 HOW GOOD ARE CHILDREN AT DRAWING CONCLUSIONS BY MAKING GENERALIZATIONS?

Description of results includes the specific features of making generalizations within each task, the commonalities found across tasks and the coherence and levels of performance in this 'process'.

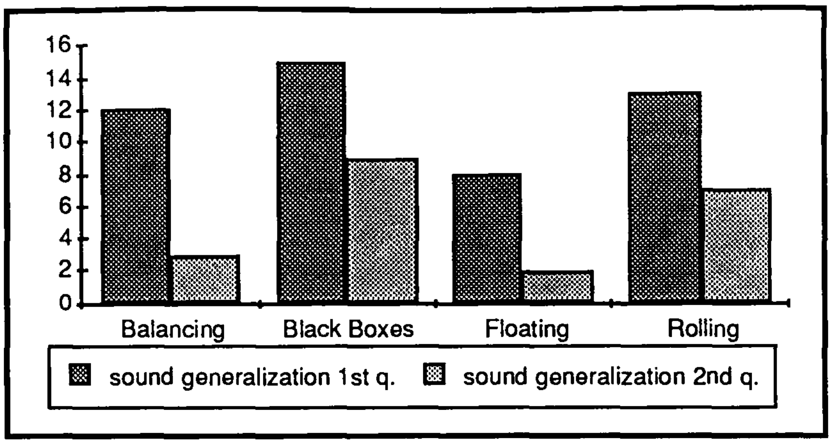


Figure 8.2.2.1: number of children making generalizations in the 1st question and number of children making generalizations in the 2nd question per task.

Balancing:

When children make a sound generalization, they tend to do it in the first question (12/24) rather in the second (3/24); making a sound generalization in the second question seems much harder (see Figure 8.2.2.1). Most of the sound generalizations consist in matching weight on either side of the balance, under the assumption of equal distances - something that they do not mention - and few take into account distance and compensate accordingly (see Figure 8.2.2.2 and Section 8.1).

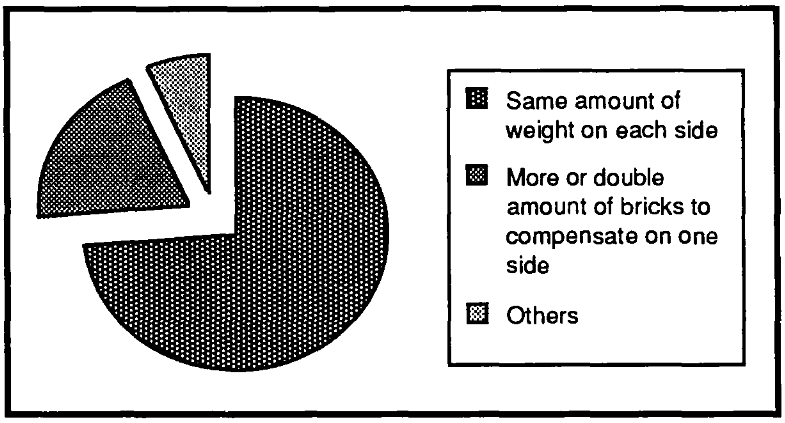


Figure 8.2.2.2: fraction of sound generalizations according to type in the Balancing task.

If one looks at making generalizations in relation to 'what-if reasoning' there is a strong positive relationship (see Chapter 7). It seems that being able to make generalizations - either related to having weight at the same or at different distances - helps in predicting the weight or distance required to get the balance level and vice versa. Good performance on these two 'processes' might explain a good performance in the task, due to their logical connection.

Floating:

The difficulty of making a sound generalization in the second question is also present here: 8/24 in the first question and 2/24 in the second (see Figure 8.2.2.1). Making generalizations related to the effect of weight, assuming that density is constant - a condition not made explicit by children -, accounts for half of all generalizations made; with fewer generalizations concerning the effect of density (see Figure 8.2.2.3). This corresponds neatly to the findings about noticing phenomena and identifying variables, where it appears easier to notice or identify as a factor the effect of weight rather than the effect of density (see Section 8.2.1 and Section 8.3.1).

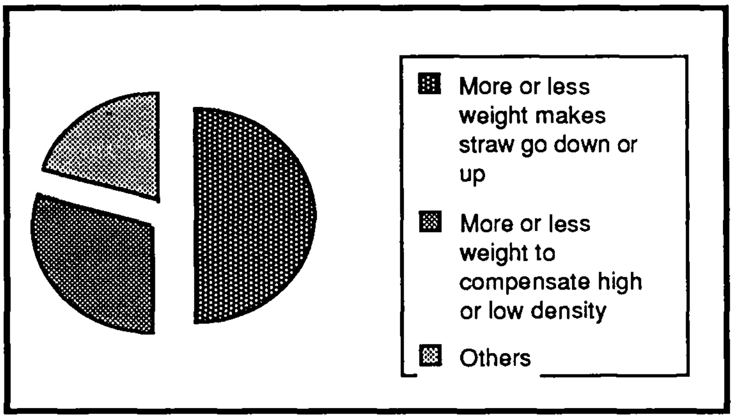


Figure 8.2.2.3: fraction of children that made certain type of sound generalizations in the Floating task.

The relationship with 'what-if reasoning', appears to be strong (as with Balancing) (see Chapter 7). Nobody who failed to make one sound generalization correctly predicted the floating of a straw.

Rolling:

Here the difficulty of making a sound generalization in the second question is less severe: 13/24 in the first question and 7/24 in the second (see Figure 8.2.2.1). Saying that the ball always goes down or the path is curved, accounts for most of the sound generalizations (see Figure 8.2.2.4). What is interesting to note is that saying that the more you pull the spring the further the ball goes or the higher the 'gun' is pointed the narrower the curve - assuming unchanged conditions of direction and velocity respectively - are infrequent ('others' in Figure 8.2.2.4). This is probably why an expected relationship with identifying variables is not in fact present (see Chapter 7).

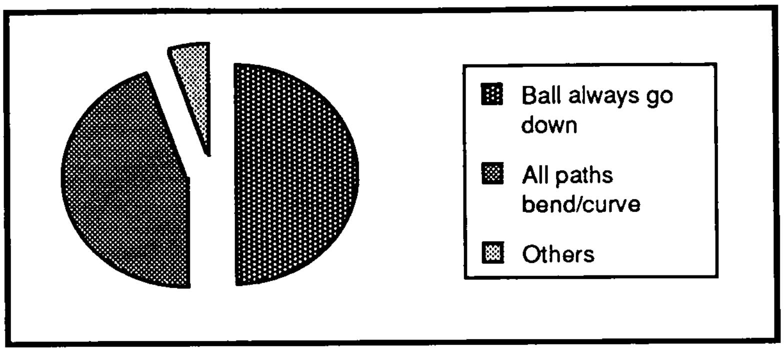


Figure 8.2.2.4: fraction of children that made certain type of sound generalizations in the Rolling task.

This result also to support the idea that children are more concerned about features of the trajectories themselves than about the connection between factors which produce them; although the way 'what-if' reasoning was represented in this task does not help to strengthen such a claim. Nevertheless, this task seems to be dominated by specific expectations (see Chapter 7).

Black Boxes:

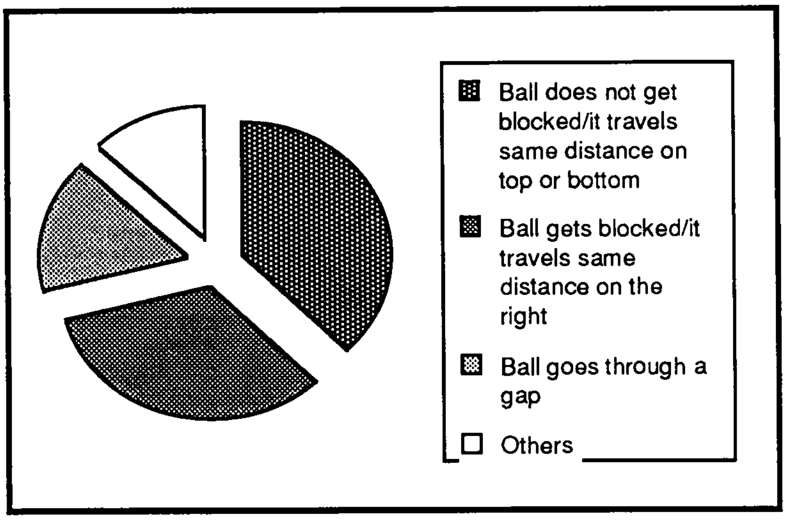


Figure 8.2.2.5: fraction of children making certain type of sound generalization in the Black Boxes task.

This is the task where more children make a sound generalization in the second question, as well as showing the best ratio (0.6) between both questions; 15/24 in the first question and (9/24)in the second (see Figure 8.2.2.1). Two kinds of generalizations, in equal proportions, account for most of the sound generalizations: that the ball gets blocked at the same distance on the right hand side or that the ball never gets blocked at the top or bottom (see Figure 8.2.2.5). These two types of generalizations are present in 'what-if reasoning situations, thus a stronger relationship was expected with 'what-if' reasoning; it turned out to be positive but only weak (see Chapter 7). It seems more difficult to see what is the same across boxes by oneself, than to discriminate boxes where the discriminative test is already set in the task; this contributes to the idea that it is difficult to see across tasks analytically (see section 8.1 and section on 'what-if' reasoning).

Commonalities :

Performance on this 'process' shows that children can mostly make one relevant generalization from their observations; in all cases making one in the second question proves to be much harder.

A strong relationship between making generalizations or drawing conclusions and 'what-if reasoning' might be expected - as in the case of Balancing and Floating -, if strong expectations or unclear identification of variables do not intervene - as they do in the case of Rolling and Black Boxes respectively. Making a sound generalization is to put in general form the relationship between variables - given certain general conditions that children do not tend to make explicit; explicable because of the difficulty in using conditionals (APU, 1988) - and 'what-if' reasoning is to predict the outcome of a particular set of conditions of such a general form.

Coherence and levels of performance :

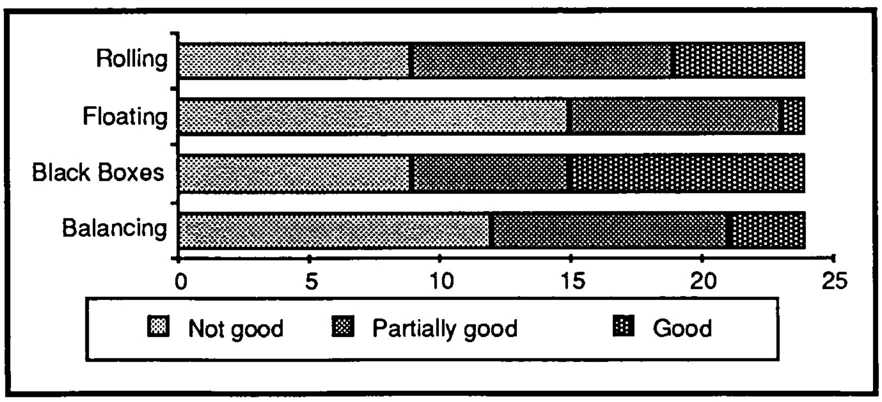


Figure 8.2.2.6: number of children scoring in Making Generalizations depending on level of performance and task.

In the Black Boxes task there is a relationship between the two questions in the sense that no child succeeds in making a second sound generalization if they fail to make a first (see Chapter 7). The small number of children making a second sound generalization reduces the possibility of detecting such relationships. In fact Black Boxes shows the highest number of sound generalizations in both questions (see Figure 8.2.2.1). Rolling follows behind Black Boxes in this respect; it shows signs of a positive relationship (see Chapter 7), but the marginal totals make it difficult to establish.

If 'good' and 'partially good' performances are taken together, Balancing and Floating appear to be the most difficult tasks in making sound generalizations, with Black Boxes and Rolling having lower and similar levels of difficulty (see Figure 8.2.2.6).

8.3 HOW WELL DO CHILDREN IDENTIFY AND CONTROL VARIABLES? HOW DOES THIS RELATE TO THEIR UNDERSTANDING OF THE TASK?

8.3.1 HOW GOOD ARE CHILDREN AT IDENTIFYING VARIABLES?

The answer to this question will include several aspects: what counts as a variable for children, what is the nature of the variables involved in the tasks, how well children just identify those variables, how well they justify them and how they perform in bringing identification and justification together.

8.3.1.1 What would count as a variable?

It is not obvious what would count as a variable in the 'production' of a specific phenomenon, despite the neat way variables are dealt with when the results of an experiment are reported in science. Much of this has to do with the understanding of what is involved in each particular task; that is, the particular understanding that children have in relation to some phenomenon. For example, in the case of the Floating task - structured version - children are asked 'what makes the straws float differently?' and some children just look at one factor, either 'weight' or 'kind of water', and others think that the 'amount of water' has something to do with it (see Figure 8.3.1.1). In this case children's understanding of what may affect the floating level of a straw, amount of water, plays an important role in the identification of what would count as a variable; in fact when children are asked to identify what the floating level depends on (identifying variables section of the instrument), 5/24 children mention amount or level of water as a factor affecting it (counted as not plausible).

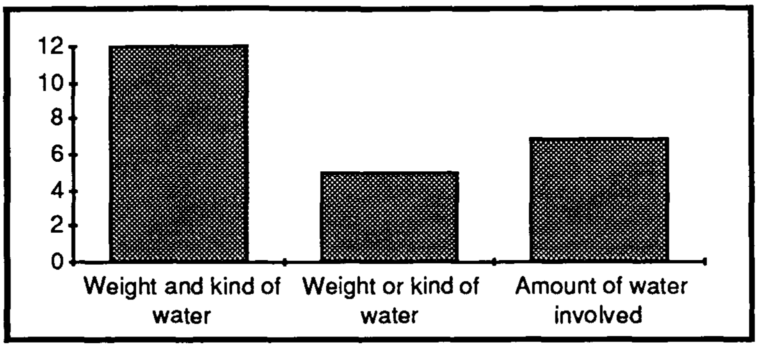


Figure 8.3.1.1: number of children answering 'what do you think makes the the straws float differently?'.

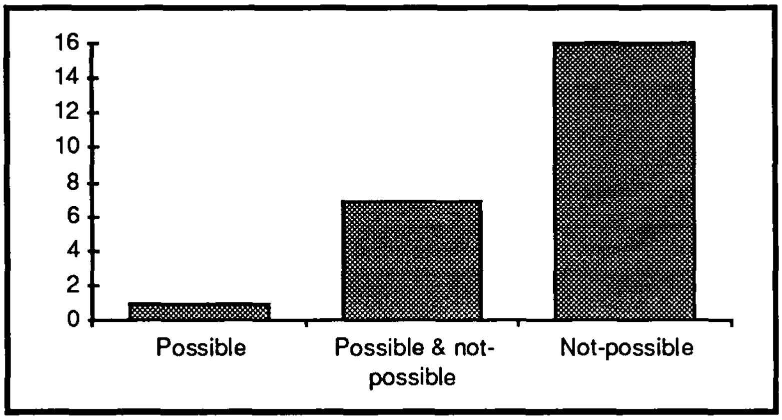


Figure 8.3.1.2: number of children that identified some trajectories of the ball - 'dropping', 'straight' or 'parabola' - as possible.

There is another task, Rolling, where what is seen as a variable seems also to be affected by expectations. In the structured version of Rolling, children are presented with three different kind of paths - 'dropping', 'parabola' and 'straight' - and asked 'which of the following paths are possible?'. The majority of children choose either impossible ones like 'dropping' and 'straight' or a combination of possible and not possible (see Figure 8.3.1.2). This suggests that in choosing 'dropping' and 'straight' the factor called direction seems not to be very relevant. The important thing seems to be an expectation that the ball would drop when losing 'force'; and not a specific path determined by a direction and a specific velocity. In fact in the investigation plus goal task children often use a lot of 'power' to compensate the slant of the board. It can be seen in Table 5, Appendix D, that the most favoured throws are 'downwards' and 'horizontal at bottom' with 'upwards' as the less favoured one. In both cases children prefer positions where the question of direction is less decisive because it can be easily countered by using more velocity. Furthermore, children identify - without justification - velocity 22 times and direction only 10 times (see section 8.3.1.2).

The evidence presented for these two tasks suggests that the understanding of phenomena can affect what a relevant factor affecting such phenomena might be.

8.3.1.2 What is the nature of a variable?

In natural sciences it is rather usual to expect rational values for the variables that are being manipulated in an experiment, to plot them and describe their relationship by a mathematical function. Is it the same case in primary science education?

One might think that a concept like weight, seen as a variable in an experiment, should be easily seen as taking fractional values as well as integer values. When answers from the written questions were classified it was found that a 'definition' of what was acting as a variable should be understood in their own terms. Children were questioned in the Structured version of Balancing, after being asked to get the balance level in different ways, 'what is one thing that getting the scale to balance depends on?' The answers given are far from saying 'the amount of mass' or 'the amount of plastic' present on each side of the scale. Some of them would say *'the amounts of brick you put on each side'*. The same child then might say to 'what difference does it make?': *'if you put the equal amounts of brick on each side it wall* [will]*balance tother* [together]*but if you put more on one side and less on the other it wall* [will]*tilt'*. From the justification given it can be seen that he has a good reason for why bricks would count as a variable or factor; that is, he knows the effect of manipulating such a variable. Certainly this child is not using a precise definition of what would count as variable, but nevertheless it is clear that it is reasonable to accept his answers as identification and justification of such variable.

Support for the claim that answers are far from saying 'the amount of mass' or 'the amount of plastic' comes from results on 'what do you think makes the balance level [in terms of weight]?' in the Structured version of Balancing. Children mostly identify what can be considered as a discrete factor (amount of dots on top of the bricks) or a combination of a 'discrete' and 'qualitative' factors (bricks having the same size), and very few identify any 'continuous' factor (the same amount of plastic) `(see Figure 8.3.1.3).

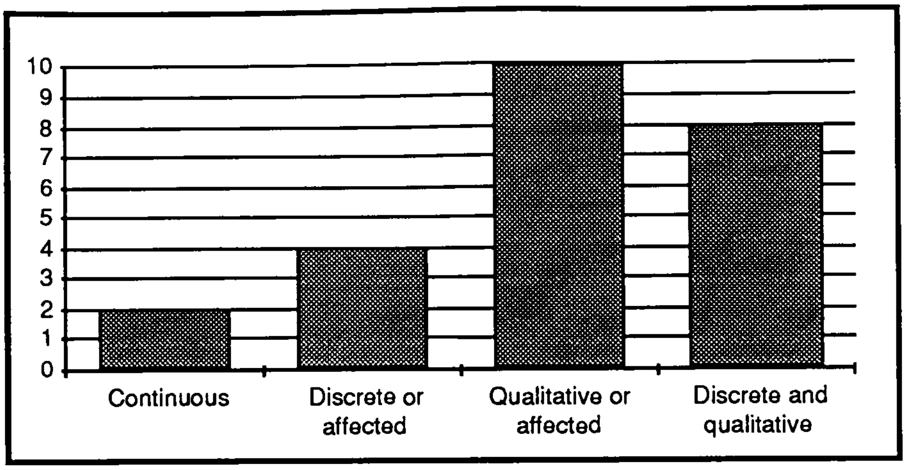


Figure 8.3.1.3: children's answers to the question 'what do you think makes the balance level?'.

There is another important element that seems to be part of the way children identify a variable; this is action. In all three cases, Balancing, Floating and Rolling, the variable that is most identified in each case is related to an action. In Balancing 'weight' is identified more than 'distance', in Floating 'weight' more than 'density' and in Rolling 'velocity' more than 'direction' (see Figure 8.3.1.4). While physical manipulation of weight in the 'Balancing' and 'Floating' tasks is possible, 'distance' and 'density' are given and therefore they cannot be manipulated in any strict sense; they only can be used or chosen. In the case of Rolling the variable that 'naturally' can be seen as related to action is 'velocity' because the effect of pulling hard or less hard is almost self evident, while with 'direction' its effect is not so evident.

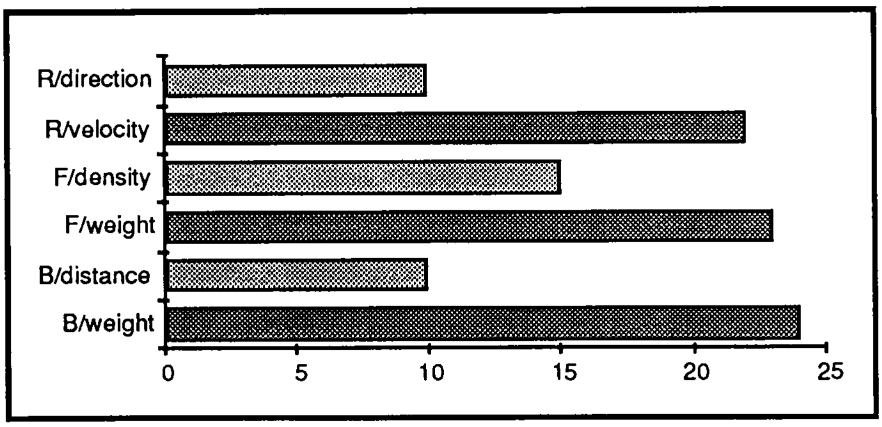


Figure 8.3.1.4: number of children that identify a specific variable in each task.

8.3.1.3 How well do children identify variables?

Children were asked twice in all structured versions of the tasks, what is one thing that the production of some phenomenon depends on. In all three cases, Balancing, Floating and Rolling, almost all identify a variable in the first question (see Figure 8.3.1.5). But there is a significant drop in each task in identifying a second variable; showing its difficulty.

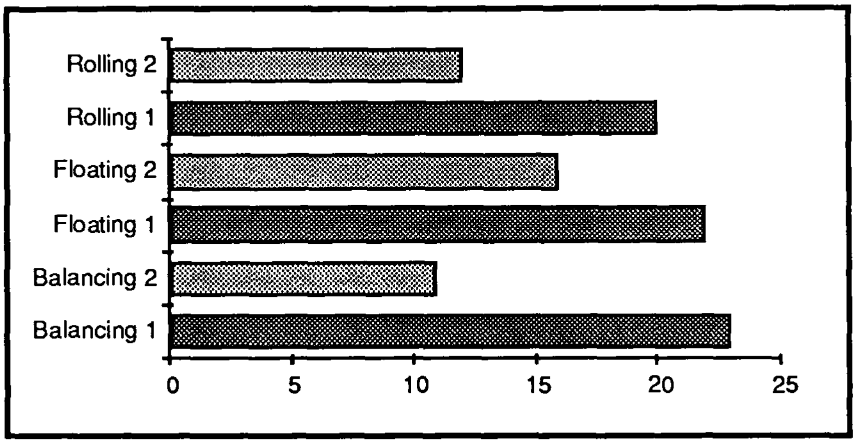


Figure 8.3.1.5: number of children that identify a variable in first and second places.

Variables related to action seem to be spontaneously produced, in that factors with such a feature were mostly identified in the first question (see Figure 8.3.1.6).

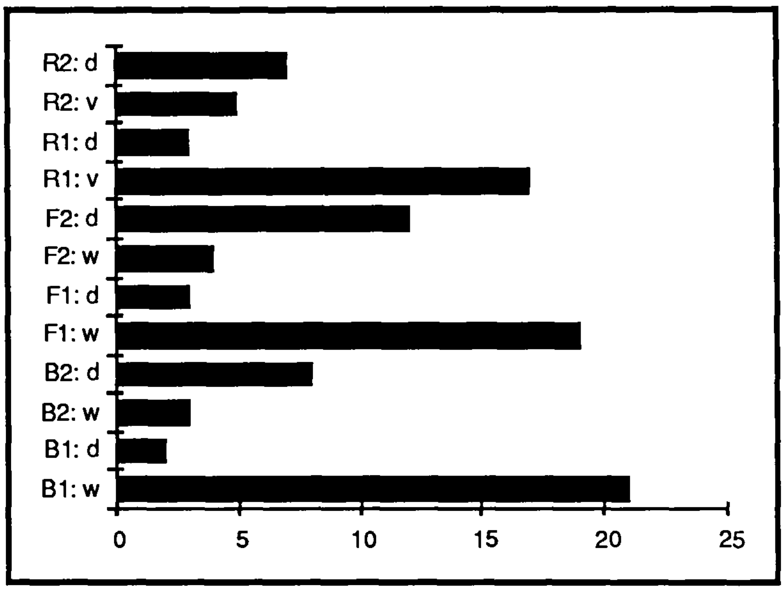


Figure 8.3.1.6: number of children identifying a particular variable in 1st & 2nd instance per task.

8.3.1.4 How well do children justify what they identify as a variable?

An identification of a variable without being able to establish the difference it makes when manipulated is of limited use. Therefore identification and justification should be considered together.

Justification of what has been identified as variable it is always more difficult than its identification (see Figure 8.3.1.7) though this difference is less evident for variables identified in the first question. In Rolling, justification is particularly difficult. Nobody who identified direction could justify it.

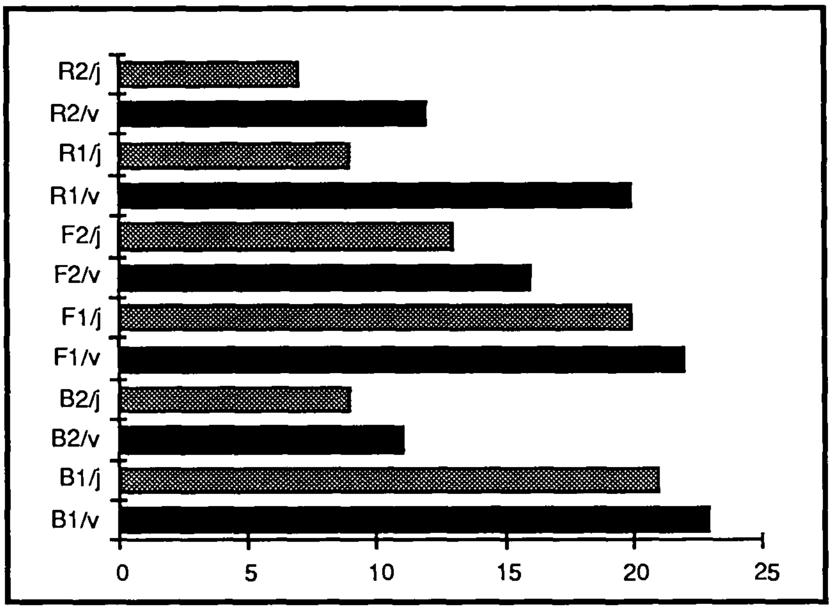


Figure 8.3.1.7: number of children per task and instance that identify a variable and justify it.

8.3.1.5 How well do children perform in identifying variables?

As explained in Chapter 5, one point was given only when both identification and justification were satisfactory (Black Boxes is excluded because no justification was asked for). The pattern is the same as before: identification plus justification of a variable drops from the first question to the second in all the tasks (see Figure 8.3.1.8). The difference is most noticeable in Balancing, with 'distance' being difficult to justify (see Sections 8.1 and 8.2.1). As mentioned above, giving any justification at all is difficult for Rolling. This may explain why children in the 'Investigation plus goal' version of the tasks, mostly try equal distances in Balancing and in Rolling mainly try to get the ball inside the 'trap' by modifying 'velocity' rather than 'direction' (see Section 8.1 and Table 6, Appendix H). As for Floating, where the role of the water is not clear (see Section 8.3.1.1 and Section 8.5), justification of density is more difficult than weight.

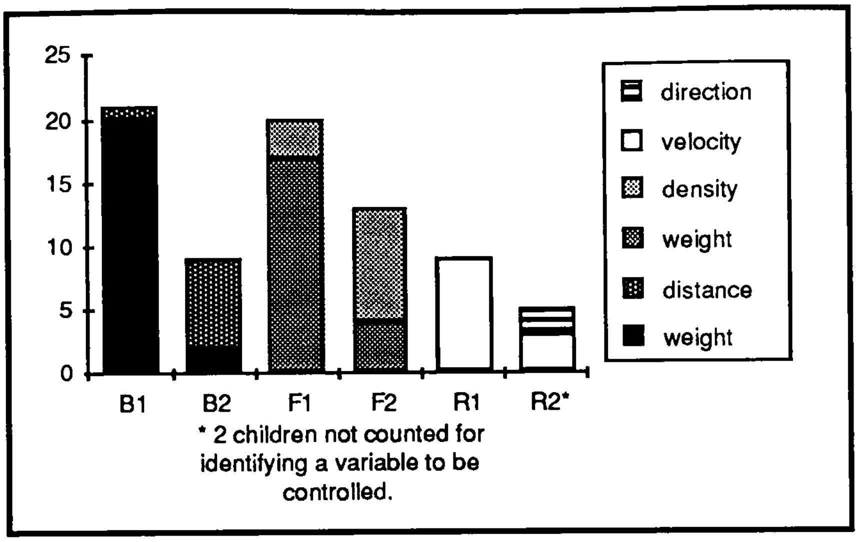


Figure 8.3.1.8: number of children identifying a variable and giving a sound justification per task and question.

In the case of Floating there is a strong relationship of 'identifying variables' with 'what-if reasoning' (see Chapter 7); meaning that identification of 'weight' and 'density' as variables helps to predict the floating level of a straw, as one might expect. In the case of Balancing, because of their different levels of discrimination, only a weak but positive relationship was found (see Chapter 7).

Commonalities :

It seems that expectations play a role in considering what may count as a variable. Possible accounts of this role were given for Floating and Rolling in the light of the evidence available. One might wonder if the same would have been found for Balancing, had children been asked about it. One might find that 'distance' is not seen so often or easily as a variable because of a dominant expectation that a balance is for equal amounts of weight on either side. The tasks considered here - Balancing, Floating and Rolling - coincide in spontaneously prompting variables connected with action. Balancing and Floating look rather similar in nature.

Coherence and levels of performance :

If the two questions related to Identifying variables are considered together, at face value, an overall performance can be calculated. In the case of Floating there looks to be a relationship between both questions (see Chapter 5). No children identified a second variable (mainly density) having failed to identify a first one (mainly weight). Results of the scoring for Balancing, Floating and Rolling can be seen in Figure 8.3.1.9). The picture shows that full identification of two variables (velocity and direction) - 'good' in the picture - is harder in Rolling and easier in the case of Floating. If identification of one variable - 'partially good' in the picture - is combined with 'good', then children do rather similarly for Balancing and Floating. It is interesting to note that Yeany et al. (1986) claim that identifying variables is one of two integrated science process skills and modes of cognitive reasoning that is acquired last.

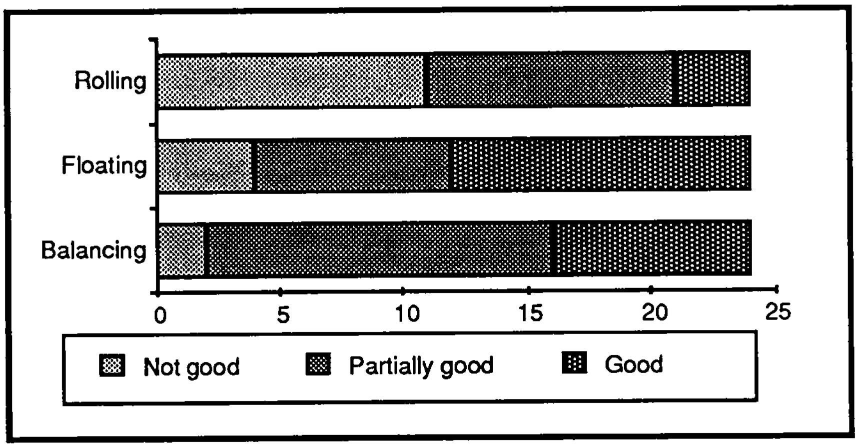


Figure 8.3.1.9: percentage of children scoring on Identifying Variables depending on band of performance and task.

8.3.2 HOW GOOD ARE CHILDREN AT CONTROLLING VARIABLES?

Controlling variables, having identified those factors that affect the problem or phenomenon under investigation, makes possible the implementation of the concept of 'a fair test'. In the simplest case of all, it can be seen as a kind of 'experimental design' where there are only two related variables - one being manipulated and the other one held constant - and the 'test' to be made 'whether it is possible - for example - to get the balance level' in such circumstances. The value of this procedure, changing systematically one factor with others constant, is of clear value for scientists in performing their experiments. This section will deal with how children as young as 11 years old manage to control the variables involved when performing the 'Investigation plus Goal' version of the tasks. The question is whether they proceed completely at random without any systematic behaviour, or whether some kind of natural or spontaneous control is used.

In order to have an idea of how systematic children being are in performing the Investigation plus Goal tasks, three ways of looking at the question were devised: at a structural level by the way children change configurations (see Chapter 5 and and Section 8.1); at a tactical level by the way children behave within a group of configurations; and an overall view.

Controlling variables when changing configuration (see Section 8.1) has an strategic value. The analytical distinction between 'configurations' defines a difference between children dealing with the same 'structural' conditions (those which would give them the same solution or family of solutions, different from other solutions in different configurations) and children trying to achieve the goal within the same configuration ('tactical' level).

8.3.2.1 How systematic are children in controlling variables when changing configuration?

Because of the differences in the size and nature of the several spaces belonging to different tasks, emphasis will be given to the particularities of each task.

In this 'structural' sense, children mostly change two variables at the same time except for the Rolling task where they mostly change just one (see Figure 8.3.2.1). Why these similarities and differences? Does it mean, because they mostly change two factors at a time, that they are not systematic? The answers resides in the nature of the tasks and, to some extent, in the way 'configurations' were defined.

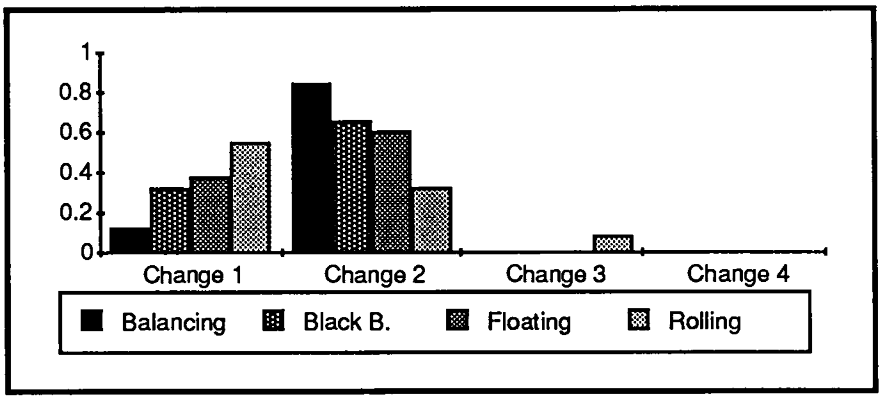


Figure 8.3.2.1: fraction of attempts made when changing configuration and changing 1, 2, 3 or 4 variables per task.

Balancing:

In this case, there is a great difference between changing one and two factors (see Figure 8.3.2.1). Because of the structure of the task and the way groups of configurations were devised here, children are obliged in this task to change 'distance' in order to change from one group of configurations (single, double, pairwise and combinatorial) to the next; this variable is what was taken as defining a change of configuration. It is clear that very few attempts, made by few children (see Tables 1 and 2, Appendix H), are made by changing just one variable: change in distance keeping the weight constant. There are two reasons that can explain this behaviour.

First of all, although it is possible in theory to keep weight constant and manipulate distance, this has some constraints. The balance has four places to put bricks on, hence if they are just currently using one brick on either side there is no way they can change to a different configuration like 'double' or 'combinatorial'; there is a minimum number of bricks to follow a strategy of keeping weight constant and change distance.

Secondly, the most easily identified variable in this task is 'weight' rather than 'distance'. Thus it is quite easy to understand that children would be more concerned with adding or taking bricks off the balance, than being 'aware' in relation to 'distance'. This situation can easily lead them to put weight where it is 'needed' - where the balance is up - and hence in the terms defined changing configuration; making more frequent the change of two variables rather than just of one. Thus it seems that children do not imagine a structure for the task other than trying the same or equivalent weights on either side using the same distance (see Sections 8.1 and 8.2).

A strategy of changing distance keeping weight constant seems not to be preferred by children (see Table 1, Appendix H). In Section 8.1 it can be seen that children spent more attempts in 'single' or 'double' configurations than in any others; this is, in symmetrical situations where different distances do not intervene. Therefore, if there is 'lack of awareness' about the role played by 'distance' (see Section 8.2) and children then search for all possible ways to get the balance level, we would see the observed pattern of behaviour.

Black Boxes:

This task presents different reasons why children mostly change two factors instead of one when changing configuration, which here means changing to a new box. There are cases of changing one 'factor' (box), keeping the test constant (see Figure 8.3.2.1 and Table 3, Appendix H), but nevertheless the relative majority of the attempts were made by changing both: box and test. It might be that, analytically, the best thing to do in order to solve the problem is to change box maintaining the same test . But children appear to follow a different way to solve the problem: they seem to try specific 'tests' for each box (see Section 8.1). In fact there are rather few changes of box - only a few more than the number required to try all boxes (see Section 8.1). The great majority of attempts are made within the same box or configuration (see Table 3, Appendix H). That children change box and test suggests that they are more concerned with finding out what is in each box rather than comparing boxes. Again we see an effect of how children imagine the structure of the task.

Floating:

In this task, changing configuration is to change beakers (density). Children mostly change weight at the same time (see Figure 8.3.2.1 and Tables 4 and 5, Appendix H). This suggests that they do not see as useful to try the last weight in a new beaker. Again, this looks like doing the problem 'object by object', as in the case of Black Boxes.

Rolling:

The behaviour in this task is different. Here children do not tend to change many things at a time, even though the opportunity to do so exists. They mostly tend to change one variable (see Figure 8.3.2.1 and Tables 6 and 7, Appendix H). This means that they change either the position of the 'gun' - up or down - or something related to the 'trap' - up or down; left or right; facing or upwards - but just one change.

8.3.2.2 How systematic are children in controlling variables within the same configuration?

Changing variables within the same configuration has a tactical value. It helps, by imposing some structure on the task, in finding the solution within a limited family of solutions.

In the way configurations are defined, in three tasks - Balancing, Black Boxes and Floating - staying in the same configuration means changing only one or no variable (see Figure 8.3.2.2). In Rolling it is possible to change up to two factors - direction and velocity. The option to change nothing is an important one that requires a separate look later on. Its importance is in replicating the results obtained in the previous attempt.

Balancing:

The variable most often changed here is weight more than distance (see Table 1, Appendix H and Figure 8.3.2.2). In this task, to change distance but not weight in the same configuration is to test the symmetrical or 'twin' position of the same configuration; testing in fact if symmetrical distances behave in the same way. To do this is to test the apparatus. Thus changing weight and distance (same configuration) have behind them two different purposes. One tries to map part of the space or set of solutions within the same configuration, while the other, tests the device itself. The strong tendency seems to be to take the device for granted, and not test it.

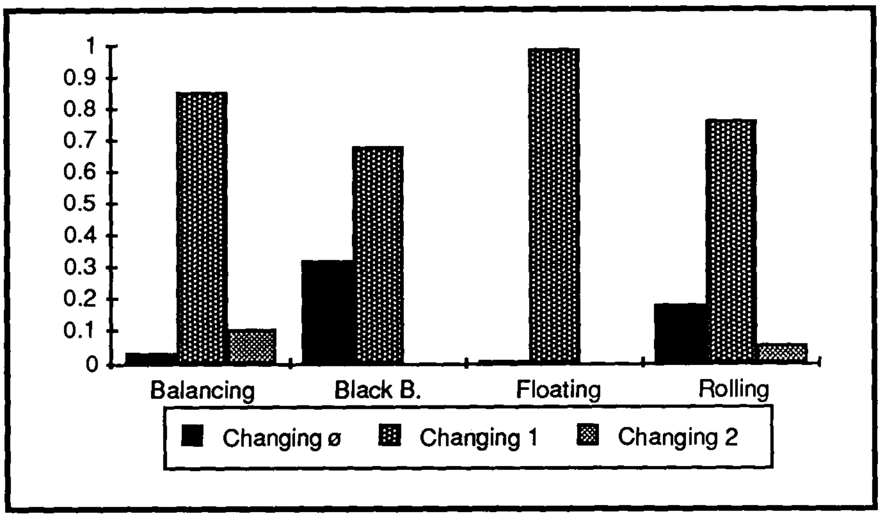


Figure 8.3.2.2: fraction of attempts made within the same configuration changing Ø, 1 or 2 variables per task.

Black Boxes:

In this task children can either repeat a test, changing nothing, or try a different test (in the same configuration). Figure 8.3.2.2 and Table 3, Appendix H shows that children mostly change one factor, but with a good proportion of repeated tests. This corresponds with the most common strategy (see Section 8.1) of taking boxes as objects - without seeing connections between them - and trying to find out which box is which by searching for the pattern within each box. Repeating tests makes sense in the nature of the task, where children cannot see inside the box.

Floating:

Staying in the same configuration here, means changing the weight (see Figure 8.3.2.2 and Tables 4 and 5, Appendix H); there is almost no repetition.

Rolling:

In the Rolling task, where they can change more than one factor in the same configuration, children mostly change one - either direction or velocity - with a moderate proportion of attempts changing nothing, and very few changing two at the same time (see Figure 8.3.2.2 and Tables 6 and 7, Appendix H). Children's behaviour looks quite sensible if one focuses on the amount of changes made: mostly one, but having the chance to do more. They mostly change velocity rather than direction (see Table 6, Appendix H); in line with what has been said before in relation to identifying variables (Section 8.2).

8.3.2.3 How systematic are children overall?

Here the intention is to give an overall picture of how children 'control variables' regardless of whether the sequence of attempts focuses on the structural or strategic level or the tactical one. For this purpose, attempts made in the same and different configurations have been added together. Additionally, to give a more complete and meaningful picture of how systematic children are, results related to whether changes were made after success or failure, and the direction the investigation takes when such changes are made will be discussed here.

From Figure 8.3.2.3 it is clear that changing one factor is a common feature for all tasks, with changing two relatively rare except for Balancing and Floating, and with very few changing nothing. This picture may suggest that children are more keen on 'exploring' the effects of one factor at a time, but are 'reluctant' to confirm what they have found.

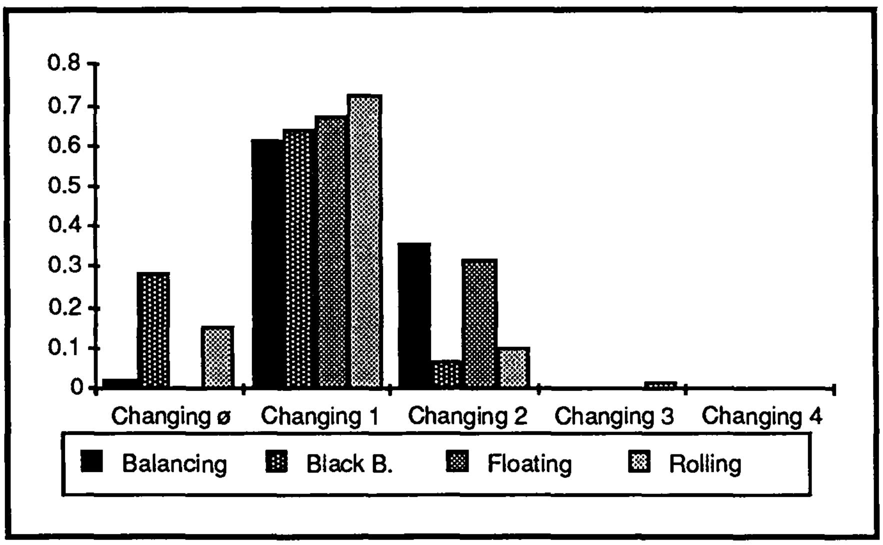


Figure 8.3.2.3: fraction of attempts made in all conditions and changing Ø, 1, 2, 3 or 4 per task.

It roughly agrees with the APU (1988) findings which show children performing well with one independent variable, but might have problems with two or more; may be those problems relate to children changing two factors at a time when changing 'configuration'.

Looking at changes of variable, depending on whether children succeed or failed, would give a clearer answer to the question of how systematic children are. Two ways of looking at systematism will be described. The first looks at changes for each task, regardless of whether attempts are made in the same or different configuration. The second looks at to what extent success and failure influence the next step in the investigation.

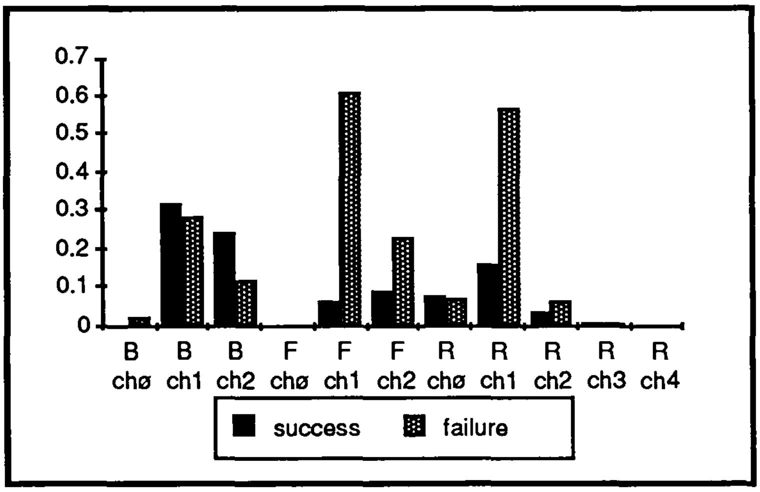


Figure 8.3.2.4: fraction of attempts made in all conditions depending on number of changes introduced and whether there is success or failure per task.

Related to the first way, Figure 8.3.2.4 shows large fractions of attempts within each task, changing one factor having failed; underlining the point made earlier that children probably know how to get the goal but not by how much, something particularly clear for Floating and Rolling. For Balancing the frequencies of attempts after success or failure having changed just one factor are similar. This might be explained by children succeeding in using same distance and changing weight to 'explore' the family of solutions, but failing when using different distances and trying to solve the problem by weight with the idea of symmetry.

Children could observe in Balancing, Floating and Rolling, the success or failure of an attempt to reach the goal. How did this influence their next step? Tables 1, 4 and 6 Appendix H show the numbers of cases of changes of variables and of configurations. These are reduced, in Tables 8.3.1 and 8.3.2 to the percentages of attempts following success, and following failure, which changed the investigation substantially or not. One way to look at this is the percentage of attempts which go to a new configuration, as opposed to staying in the same configuration (Table 8.3.1). Another is the percentage of attempts changing 2 or more variables, as opposed to changing one or none (Table 8.3.2). The two are almost, but not quite, the same thing.

	Balancing	Floating	Rolling
After success	37%	90%	27%
After failure	28%	46%	12%

Table 8.3.1: percentage of attempts changing configuration as opposed to staying in the same configuration, depending on success or failure, for all tasks except Black Boxes.

	Balancing	Floating	Rolling
After success	43%	58%	16%
After failure	27%	27%	10%

Table 8.3.2: percentage of attempts changing 2 or more variables, as opposed to changing Ø, 1, depending on success or failure, for all tasks except Black Boxes.

In every case, the percentage of attempts which change the investigation substantially, is larger following success than failure, on either measure. This is not to say that they are all the same. In Floating, changes to a new configuration (new beaker) are naturally very common indeed, after success. In Balancing, and Rolling, there is a much greater tendency to continue exploring 'nearby' configurations. but the evidence here is that children do to some extent make bigger changes to their investigation following successes. This might be interpreted as them treating such a task (Investigation plus Goal) as a number of subtasks. Note that had we defined 'success' in Black boxes as deciding about a box and starting on another, 100% of attempts after success would - by definition - have been changing configuration.

8.3.2.4 How well do children control variables and replicate results?

An attempt was made to devise a score for being systematic in controlling variables. Basically, children were more rewarded if they changed one variable at a time instead of two, regardless if that was after success or failure, but changing two after success had some reward while two after failure had none. The case of changing nothing after success, this is replicating, was scored separately (Black Boxes it is not included here because of not having outcomes as success or failure).

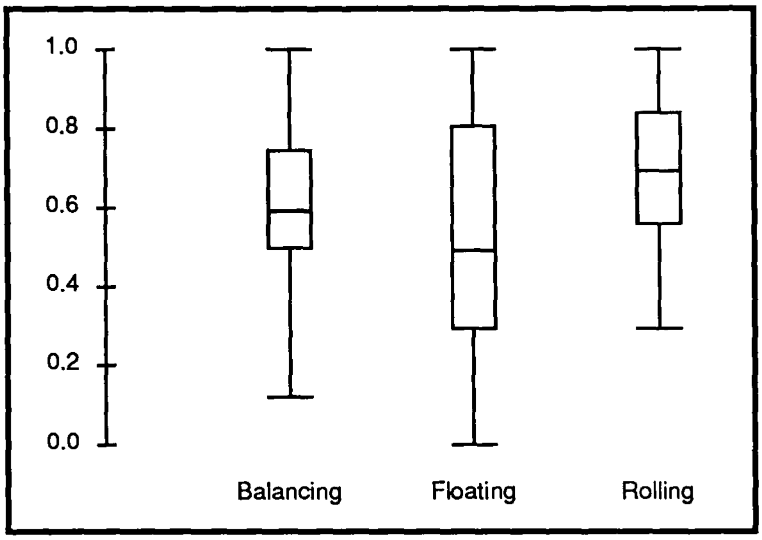


Figure 8.3.2.5: level of systematism in controlling variables when changing configurations per task.

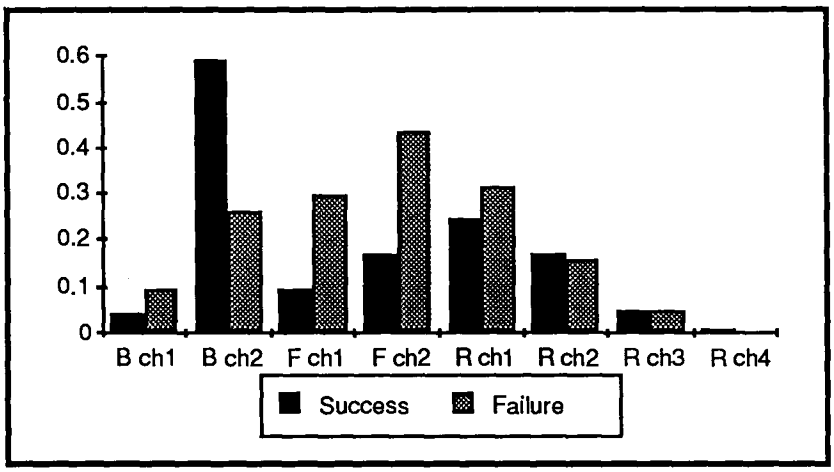


Figure 8.3.2.6: fraction of attempts made changing configuration and changing 1, 2, 3 or 4 variables after success or failure per task.

The results suggest that children are less systematic in Floating and more so in Rolling, with Balancing between them (see Figure 8.3.2.5), when changing configurations. Floating may come out low because of children's difficulty of judging floating levels, with a relatively large fraction of attempts changing two factors after a failure (see Figure 8.3.2.6). Rolling looks rather systematic, although the search is restricted to some group of configurations rather than others (see Section 8.1). Almost the same could be said in relation to Balancing, though there is a lack of attempts at difficult configurations. Figure 8.3.2.5 shows that some children are able to perform at the highest level of being systematic, as it has been scored, with children being systematic between 50% and 70% of the time. These results suggest that to be systematic when changing configuration is not as easy as one might could think. Scores for being systematic when dealing with the same configuration prove to be almost meaningless given the restrictions imposed by the way configuration was defined.

In relation to 'replicating' or 'checking' results, what was scored was the presence or not of no changes after success in relation to the total number of attempts made after succeeding, when trying within the same configurations. In the case of Black boxes what was rewarded was no change of test when dealing with the same box. What this procedure shows is that only two tasks, Black Boxes and Rolling, present some signs of replication (see Figure 8.3.2.7). The other two, Balancing and Floating, show only traces of it, with some outstanding values. When 'replicating' is present it is probably due more to the nature of the tasks, where repeating is important either to be sure where the ball is running - Black Boxes - or to make sure - and a little bit of fun - that it is possible to get the ball inside of the trap more than once (Rolling).

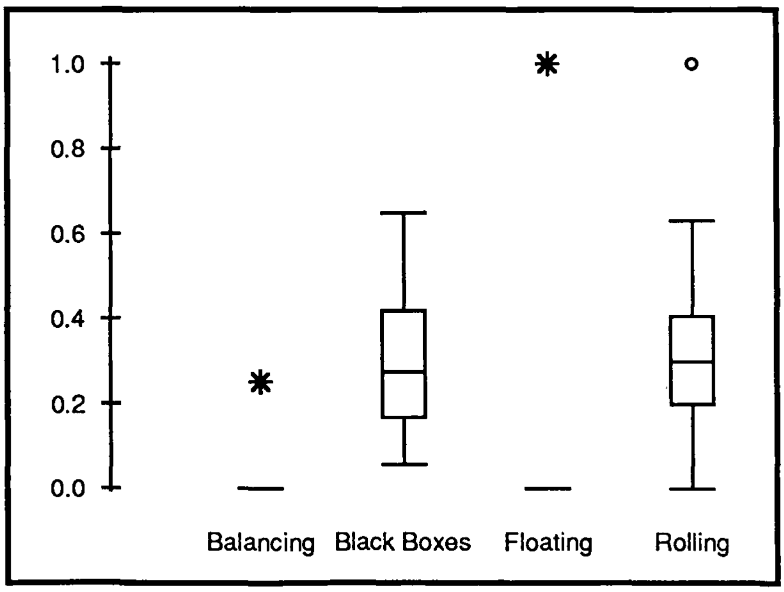


Figure 8.3.2.7: level of replicating results per task.

In summary, it seems as if children do not tend to replicate their results - finding out about the 'mechanisms' responsible for such behaviours - and they use most of the attempts in trying to get them right - to achieve the goal for each task without knowing exactly by how much -, suggesting that children's behaviour appears to be led by a desire of success more than by a desire to get knowledge from the task that is being performed. The APU (1988) also reports children's lack of checking on their results.

8.4. HOW GOOD ARE CHILDREN IN 'WHAT-IF' REASONING?

To know how phenomena happen in science, it is necessary to establish the link between the variables involved and to be able to predict the outcome, given certain conditions. Such links and forecasts are looked at in the form of 'what-if' reasoning.

As indicated in Chapter 7, it was possible to devise very similar questions to elicit 'what-if' reasoning for Balancing and Floating. Results for Rolling are less useful in terms of 'what-if' reasoning, but confirm other findings. Results for Black Boxes, although within the concept of 'what-if' reasoning, place more emphasis on discrimination (called for this reason hypothetical reasoning: hr) rather than prediction; less strictly related to 'what-if' reasoning but still pertinent are setting conditions practically (p) and devising a test (t).

Balancing:

Less than half of the students (11/24) compensate for a longer distance by predicting less weight needed to get the balance level (see Figure 8.4.1). Fewer still (7/24) can give a plausible justification (see Figure 8.4.2) of such behaviour, showing a knowledge of the precise relationship between the variables. On the other hand, more than half (16/24) take into account distance in order to compensate a big weight with a short distance (see Figure 8.4.1) and 12/24 give a plausible reason (see Figure 8.4.2).

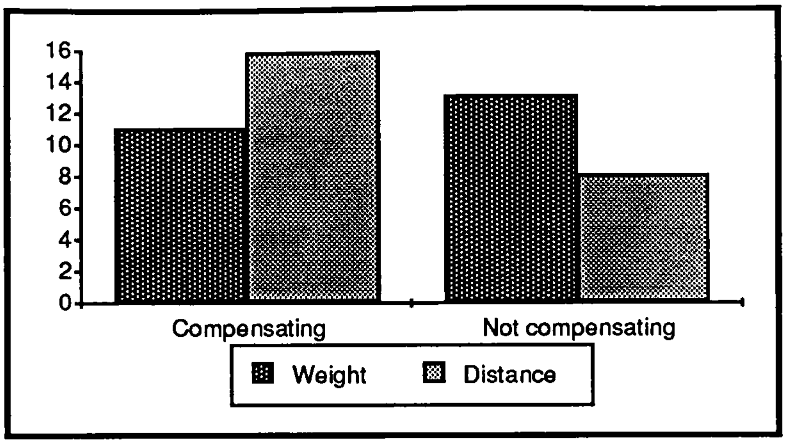


Figure 8.4.1: number of children compensating by weight or distance in the Balancing task.

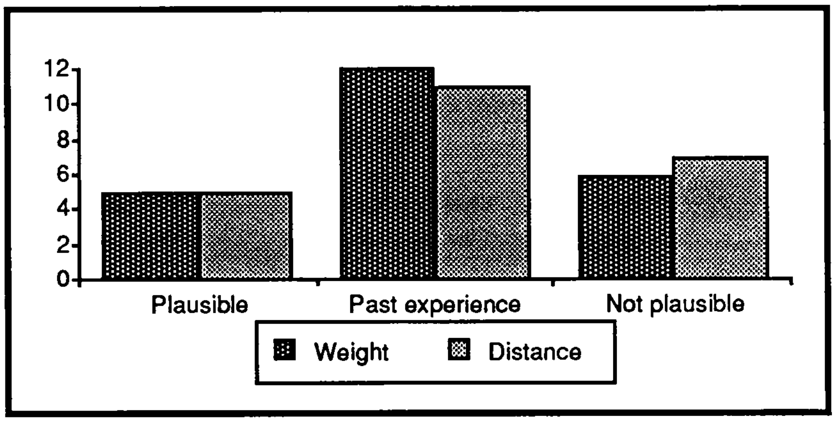


Figure 8.4.2: number of children justifying their prediction in the Balancing task.

The considerable number of children justifying their predictions using past experience as an argument indicates that children did pay attention to what was happening, but maybe without understanding how weight and distance are related.

Floating:

More than half of the students (15/24) predict which density makes the straw rise to its highest level (see Figure 8.4.3), but few (5/24) can explain why (see Figure 8.4.4). Predicting and justifying which density would sink the straw to its lowest level is similar (see Figures 8.4.3 and 8.4.4). Thus, it seems that about one third of the students appear to not understand the effect of density when the weight is constant.

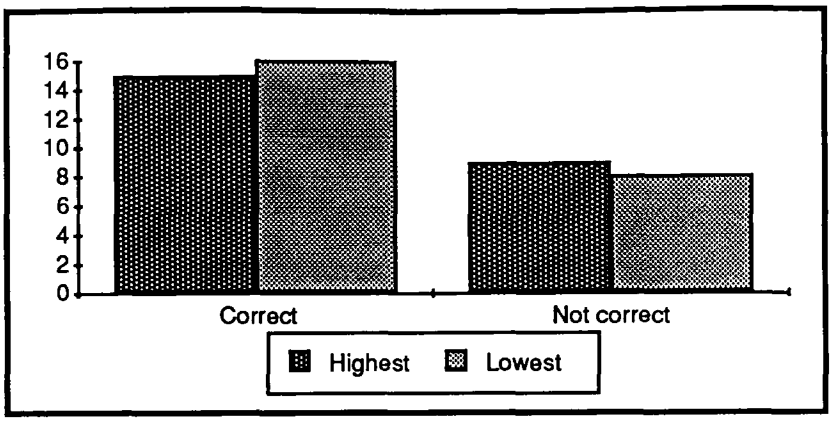


Figure 8.4.3: number of children predicting correctly the highest and the lowest floating levels in the Floating task.

The high proportion of justifications based on past experience suggest that children notice 'significant' or 'outstanding' isolated behaviours of the straws, but might not have good reasons why things happen in a particular way; making it difficult to believe that children have a clear notion of density. This claim can be supported by the difficulty they exhibit in identifying 'kind of water' as a variable (see Section 8.3.1) and the tendency to not compensate with weight when changing configuration or density (see Section 8.1).

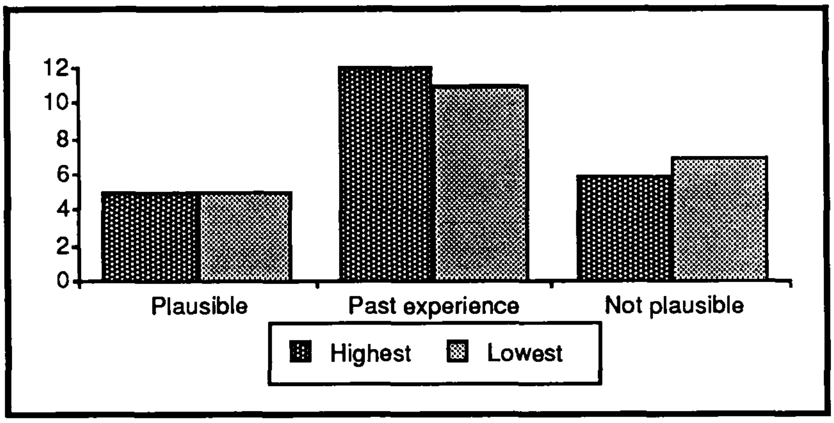


Figure 8.4.4: number of children justifying the highest and lowest floating levels in the Floating task.

Roller Ball:

In this task children set one variable - either the position of 'gun' or 'trap'- with the other one already fixed, and what they are asked to predict is not a 'value' of the variable selected (velocity or direction) but the kind of trajectory they expect from such relative positions.

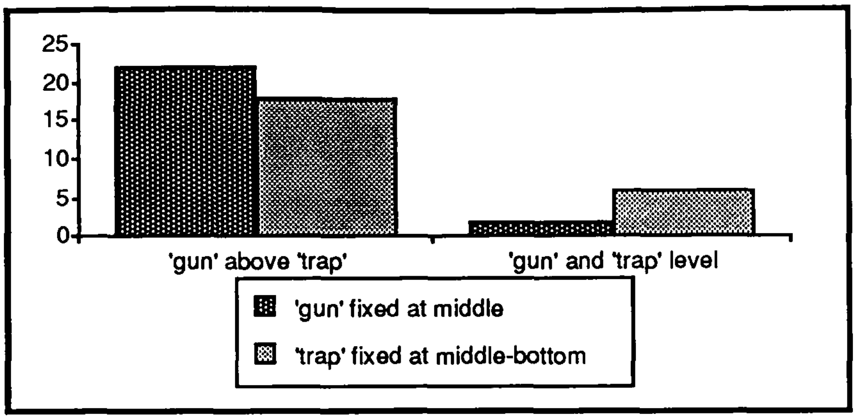


Figure 8.4.5: number of children choosing a place for 'gun' and 'trap' being fixed either the place of the 'gun' or of the 'trap' in the Rolling task.

It was found that children rather prefer to 'fire' downwards -'gun' above the 'trap' - regardless of the variable they fixed (see Figure 8.4.5). In addition, around half of the students in both situations predicted not plausible paths (see Figure 8.4.6). These results correspond to the settings for 'gun' and 'trap' that they prefer in the Investigation plus Goal version of the task (see Section 8.1) and the findings about the kind of trajectories they think are possible (see Section 8.3.1.1). It supports the previous claim (see Section 8.3.1.1) that 'firings' like 'straight' and 'dropping' - most of the impossible paths they predicted - do not need a tight control over direction when rolling the ball into the trap. Furthermore, it was found in Chapter 7 that children's predictions tend to relate to the kind of paths they expect to observe; that they are noticeably affected by expectations (see Section 8.2.1).

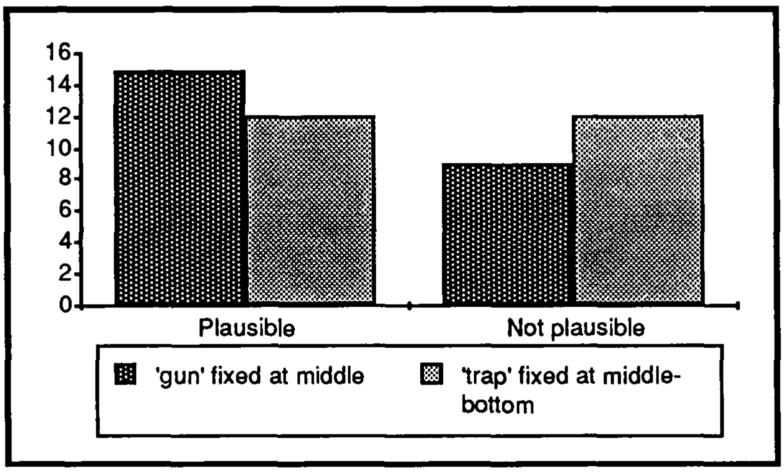


Figure 8.4.6: number of children predicting a plausible path when either the 'gun' or the 'trap' is fixed in the Rolling task.

Although justifications based on past experience are present in this task (see Figure 8.4.7), they tend to be descriptions of what happened; but are far from giving a clue about the relationship between direction and velocity.

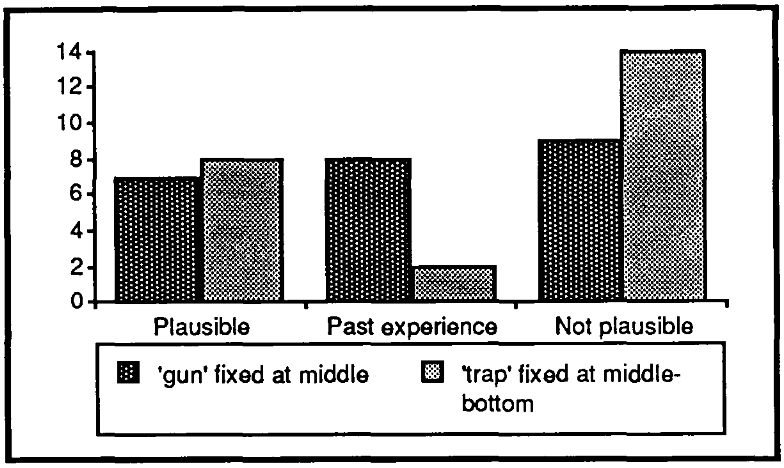


Figure 8.4.7: number of children justifying their prediction depending on whether the 'gun' or the 'trap' was already fixed in the Rolling task.

Commonalities :

One of the main commonalities of these tasks is the presence of past experience as an element of justification. However important is the presence of such element in children's behaviour, it shows at least that children pay attention to what is happening, although it does not resolve the problem of understanding the functional relationship between variables in relation to the 'production' of a particular phenomenon (e.g. balance level, required floating level, ball inside of the trap).

Another, is the difficulty in saying why something would happen given certain conditions; something that can be judged by the number of children who point out correctly the outcome but cannot say why. This is not an easy task, something already detected in the case of identifying variables (see Section 8.3.1.4). The difficulty seems to reside in the explicit knowledge and understanding of the relationship between variables it demands.

Finally, Balancing and Floating coincide in showing children having problems in knowing how to compensate the system; either using weight and distance or weight and density. This cast doubts on whether children establish a causal relationship between such variables and to their exact functional relationship. In addition, it was also found in relation to these tasks that 'predicting' or 'what-if reasoning' tends to relate to 'making generalizations' and 'identifying variables' (Chapter 7), suggesting that children's performance in predicting might be better explained by these 'processes' in these tasks; something discussed in Chapter 9.

Black Boxes:

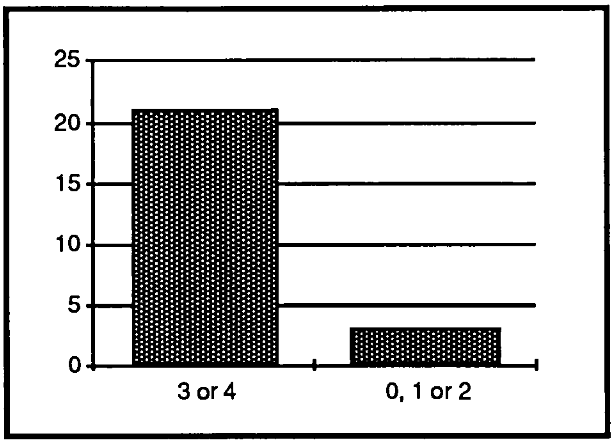


Figure 8.4.8: number of children that practically set the ball in a certain number of corners in the Black Boxes task.

Children can easily set conditions practically (t). The majority of them (21/24) are able to set the ball either in three or four corners of a box (see Figure 8.4.8). This suggests that children seem to have little problem in knowing where the ball is; the first step in finding out which box is which.

Children look perfectly capable of making judgments based on hypothetical reasoning (hr); a kind of 'what-if reasoning' used in Floating and Balancing. Generally, most of the children (18/24) make a correct or a partially correct discrimination, whether positive or negative, and can point out the lack of discrimination in a positive or negative form (see Figure 8.4.9). This suggests that children seem capable of deriving logical consequences from a hypothetical action set for them. This rather good performance seems in line with what Evans (1988) calls 'realistic content', which can facilitate logical performance; or Pollards's (1982) 'thematic content' rather than abstract.These results contrast with the ones obtained from using hypothetical reasoning in solving a problem. This is, they can derive logical consequences from a hypothetical action, but they seem to have more difficulty in using such reasoning in devising a test to find out which box is which (see Figure 8. 4.10). These results may suggest that children are logical when the elements to solve the problem are broken down in pieces and situations are already set for them, but tend to have difficulties when such logical consequences have to be derived for practical purposes by themselves.

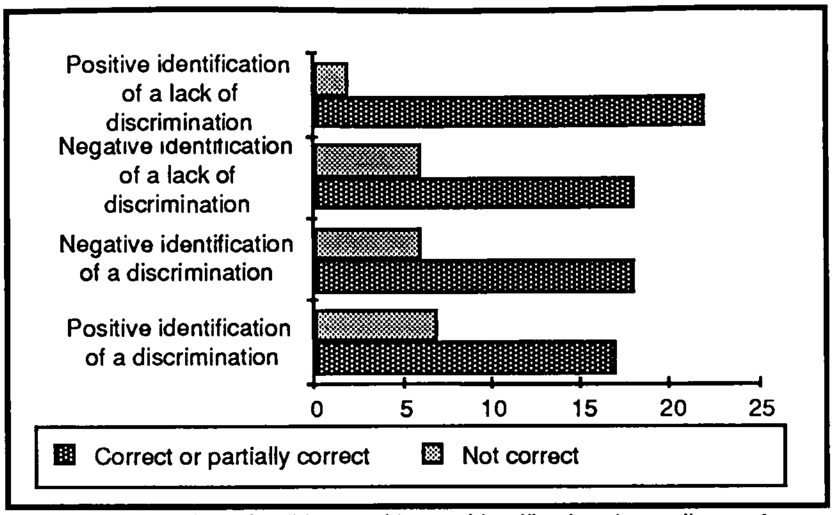


Figure 8.4.9: number of children making an identification depending on form and result in Black Boxes.

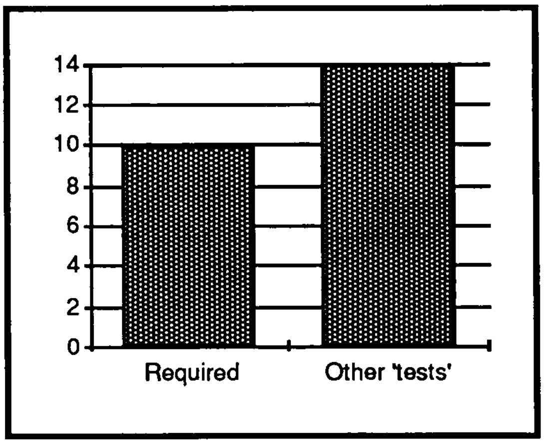


Figure 8.4.10: number of children that establish a test in order to make a discrimination of one of three boxes in the Black Boxes task.

Coherence and levels of performance :

Relationships within Balancing and Floating were found (see Chapter 5). In the case of Balancing the relationship is that essentially no children predict the correct value for weight if they fail to predict the correct value for distance (see Chapter 5), and in Floating that no children predict the correct value for the highest density if they fail to predict the correct value for the lowest.

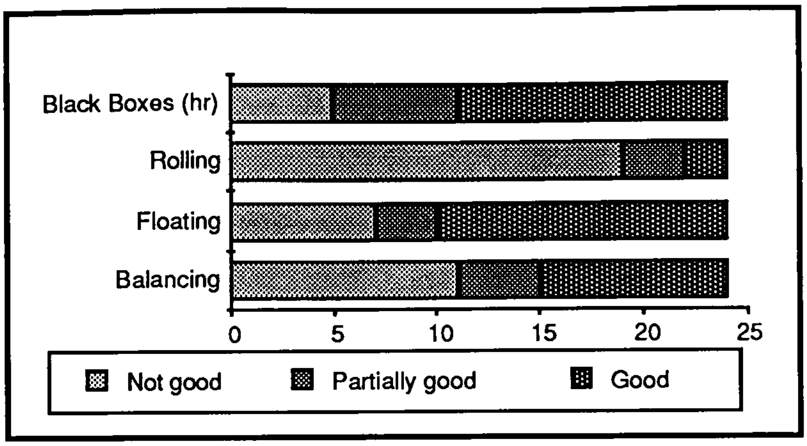


Figure 8.4.11: number of children scoring on 'What-if' Reasoning depending on band of performance and task.

In Figure 8.4.11 it can be seen that predicting is easier in Floating than in Balancing; suggesting that the understanding of the relationship between variables in Balancing is more difficult. Rolling, although different to Balancing and Floating, looks difficult - mainly by predicting non plausible paths. If 'good' (two correct predictions and justifications) and 'partially good' (one correct prediction and justification) are taken together, more than half of the students do well in Balancing and Floating. As for Black Boxes (hr), surprisingly, children did better than in Floating if 'good' (two correct discriminations) and 'partially good' (one correct discrimination) are taken together; suggesting that children are capable of at this level of abstract reasoning.

8.5 WHAT EXPLANATIONS OF THE PHENOMENA DO CHILDREN GIVE?

To explain why physical, chemical or biological phenomena (to mention just some of the phenomena science is interested in) happen as they do, is fundamental to the scientific enterprise. Imagining causes for phenomena is of vital importance in the acquisition of knowledge. To imagine a cause for a pattern of happenings like 'the more shots you add the most [more] it [the straw] sinks', is to give part of an explanation. But a causal explanation calls for more information than just a simple association of events: *if theirs* [there is] *to* [too] *much weight the boat will sink if their* [there] *isnt* [isn't] *so much weight it will float.* It should be able to describe the causal mechanism responsible for the phenomenon; something that requires in many instances the use of a larger and more general scheme that often uses 'invisible' entities - like atoms, electrons or genes.

In this section will be described how children explain some phenomena in each of the four tasks, how such explanations are related to other 'processes' and children's levels of performance.

Balancing:

Children give similar types of explanations for why the balance is tilted or level using either bricks or fingers (see Figure 8.5.1).

When the balance is tilted almost half of the students (11/24) explain the effect of bricks and fingers using arguments related to the apparatus and to the clear mismatch between both sides: '*there was nothing on the other side*' (see Figure 8.5.1). The same figure shows children (8/24) explaining the same effect in terms of bricks and fingers having the same property: heaviness, like in '*Both* [bricks and fingers] *are heavy*'. But, what does this concept mean? What does it imply? Only few of them (5/24) offer an explanation which uses an underlying mechanism: '*Both* [bricks and finger] *push down the same*' (see Figure 8.5.1).

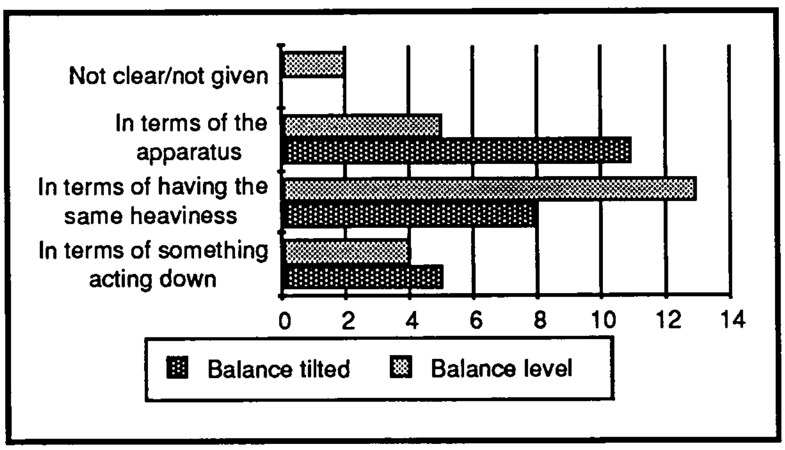


Figure 8.5.1: number of children imagining a cause for the balance being tilted or level in the Balancing task.

Although arguments based on heaviness can be seen as 'conceptual' and as 'detached' from observables, they seem to be very close to the idea that identical or equivalent amounts of bricks on both sides makes the balance level. And they say nothing in relation to the 'mechanism' attached to weight: the notion of something always acting down; that it does not matter if bricks, fingers or something else are being used.

'Imagining causes' has some relationship to 'what-if' reasoning, in that almost no child imagines a cause if they fail to predict correctly (see Chapter 7). This result makes perfect sense - if you do not know what will happen you can hardly imagine why it happens.

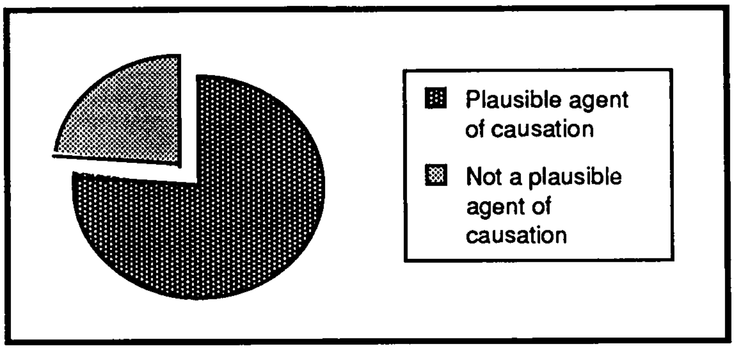


Figure 8.5.2: proportion of children that identified an agent of causation in the Black Boxes task.

Black Boxes:

The great majority of children can identify a plausible cause for why the ball following different paths inside the boxes: *'they* [boxes] *got different patterns'* or *'the walls are at different places'* (see Figure 8.5.2). They probably rely on the information given in the pictures, and on their experience of moving the ball around the boxes. It certainly seems too much to expect explanations based on physical laws of movement.

Floating:

Just above half of the children do not give water a role to play in keeping the straw up (see Figure 8.5.3). Some of them say that it is *'because of the shots'*, some *the straw has plasticine'* [in the bottom preventing water from coming in], and some *'because the straw is made of plastic'*, but they do not give water a role to play. Those who do might say that *'thick water like E will keep the straw above water'*, but others that is *' is because of the amount of water'*. Therefore, although some children imagine an interaction between the straw and the water, they do not always know the nature of the cause; some of them are identifying a 'variable' without causal power - the amount of water.

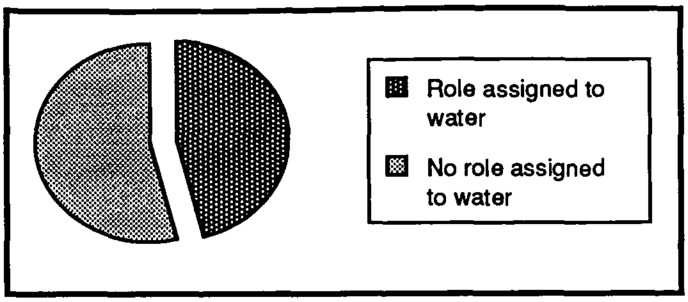


Figure 8.5.3: proportion of children that assigned or not a role to water in keeping the straw up in the Floating task.

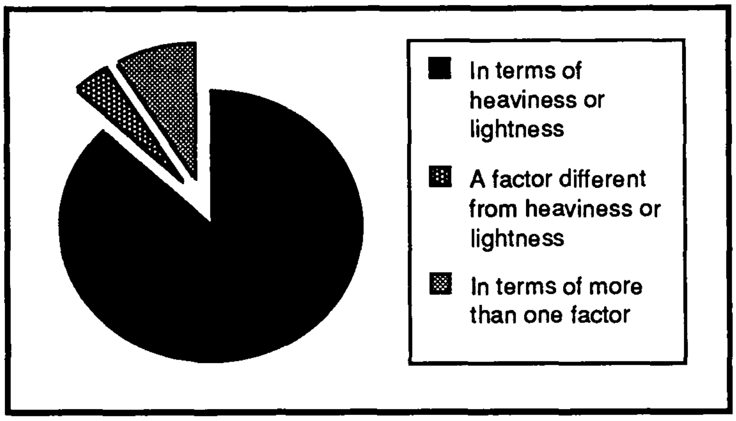


Figure 8.5.4: proportion of children giving an explanation for why things float or sink in the Floating task.

The great majority of children explain floating and sinking only in terms of how light or heavy objects are, only one of them in terms of the amount of salt in water, and two of them in terms of both object and liquid acting together - *'the weight and pressure of water'* (see Figure 8.5.4). These results support, with those already explained in the previous paragraph and the ones found in the section on Identifying variables (8.3), the claim that the water density is not very much seen as a causal factor.

When children are asked, extending the idea of a large density to the limit, if a liquid could exist that makes everything float in it, they are evenly divided (see Figure 8.5.5). The main reasons given, quite reasonably, for it being possible are based on the 'thickness', 'hardness' of the liquid or something added to it; with eight out of twelve students that said 'yes' (see Figure 8.5.5). Those who said that is not possible, mostly argue in terms of heaviness (see Figure 8.5.5); similar results were previously shown in relation to the role assigned to water (12 giving water a role to play and 12 not).

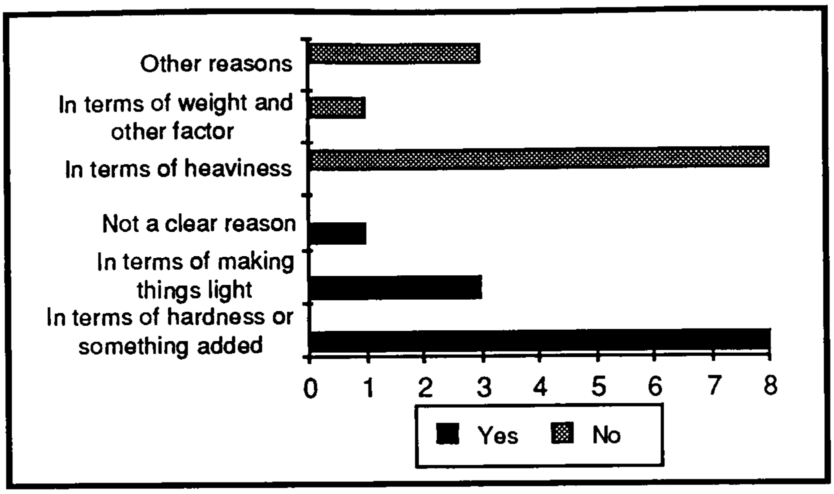


Figure 8.5.5: number of children that give a reason to why a liquid that makes everything float in it could exist according to having answered yes or not in the Floating task.

There seems to be a dichotomous explanation of floating and sinking; it is either the role of the object (weight) or the role of water ('thickness', 'hardness'), but not both.

Rolling:

Children's 'identification' of causes for two different ways of rolling the ball on the board are very similar. Explanations are mostly based on actions: *'because you pull the gun hard'* (see Figure 8.5.6). The next most preferred explanation, based on 'objectified actions', is again grounded in actions, but without a personal connotation: because of *'the way* [the gun] *is positioned and the force'* (Figure 8.5.6). It relies on an impersonal description of what they have done. Only a few identify external conditions, such as *'because it* [the board] *is tilted'*.

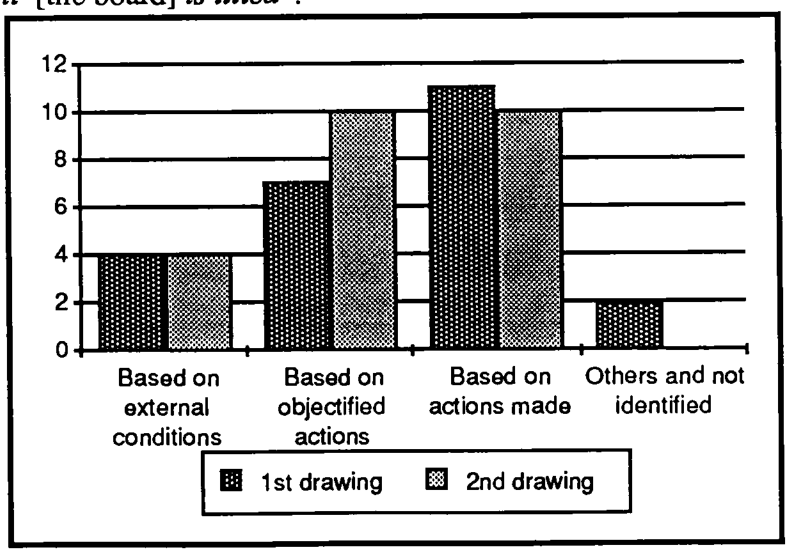


Figure 8.5.6: number of children identifying a cause for two different rollings in the Rolling task.

Commonalities :

Children's behaviour in identifying causes looks as if it is difficult for them to imagine agents of causation that are beyond the domain of the senses or action. Explanations like 'both fingers and bricks are pushing down', or things float or sink 'because of the weight and pressure of water' and others like 'because the board is tilted', are not very common. Causes related to imagination or which go beyond the senses are not easily accessible: cause is still modelled on action.

Coherence and levels of performance :

An analysis of internal coherence for the Floating task seems to show that there is a relationship between two of the three questions: giving water a role to play and the existence of a liquid that could make everything float in it (see Chapter 5).

Figure 8.5.7 shows levels of children's performance on this 'process' of imagining causes, when their answers are scored (see Chapter 5). Balancing and Rolling have high and similar levels of difficulty, both being more difficult than Floating 1&3 (role assigned to water and existence of a liquid that makes everything float in it put together) and than the Black Boxes task which is the easiest task. Apparently the more difficult one is to imagine a more complete causal explanation for why things float or sink (F2), because the majority of them seem overwhelmed by reasons based only on the heaviness or lightness of the objects and not taking both heaviness and the active role of the liquid as responsible for such phenomenon like floating and sinking.

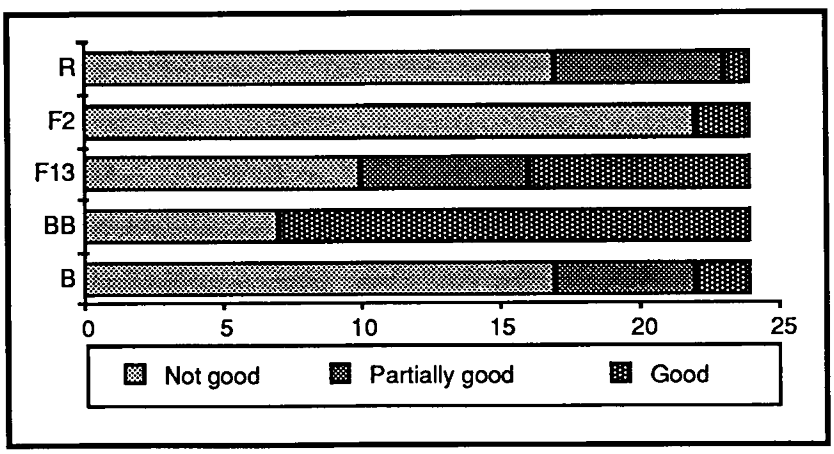


Figure 8.5.7: number of children imagining a cause depending on their performance and the task.

One possible explanation to why Balancing, Rolling and Floating (F2) look quite difficult could be that the agents of causation in all of them are not perceived directly by the senses. The apparent lack of difficulty for the Black Boxes task could reside on the strong connection between children's actions and the cause responsible for the ball following different paths.

8.6 DO CHILDREN MAKE NOTES WHEN DOING AN INVESTIGATION?

Making accurate and relevant notes has been praised in natural sciences as well as in other fields. The children making the investigations had the opportunity to write down what seemed to them interesting, striking or just adequate to their purposes. They were not obliged to do so.

Balancing:

The great majority of children did not make use of the sheet of paper provided; it was left blank (see Figure 8.6.1). The one who made relevant notes does not show a wide range of comments. The other two simply counted how many successes they got.

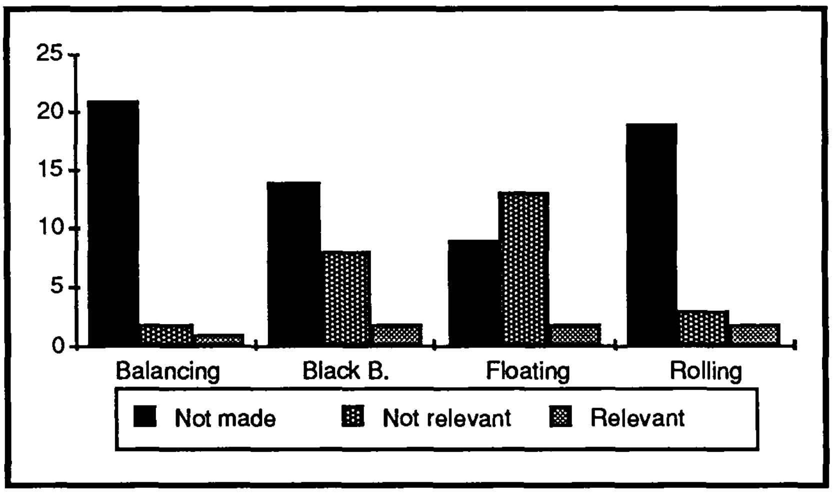


Figure 8.6.1: number of children that make notes depending on how relevant they were and the task.

Some possible types of relevant notes that could have been made are related to the use of same or different distances - with their variations in each case - and on what combinations of identical or equivalent bricks got the balance level.

There are no notes which set out a plan for the task, such as drawing possible combinations to try.

Black Boxes:

More than half of the students did not make notes (see Figure 8.6.1). One of the children drew what were the relevant features of each box: if the ball could go through or not etc. The rest of the students just drew which box got which pattern; probably to fulfil a practical requirement, that is, not to forget which box was which.

They could have made notes in relation to the outcome of tests comparing boxes, or whether certain patterns seemed similar in two different boxes; but this kind of comment is not present.

Floating:

Here the number of children not making notes is less than half (see Figure 8.6.1). Children tend to make notes related to the amount of shot needed in each beaker to get the floating mark at the desired level. One noted the shots required each of the two times the beakers were tried; probably as a checking strategy.

Children could have made notes about the effects seen - for example: changing beaker with the same weight made the floating level vary, or that they found different readings for the same beakers - depending on how they put the straw in water. Neither are present. They seem to make notes - when they do - to help their memory and be able to tell how many shots are required in any beaker.

Rolling:

Most of the children do not make notes (see Figure 8.6.1). Two wrote how many attempts were made or described a result. One of them drew 16 possible combinations of relative positions between gun and trap.

Children could have made notes dealing with the structure of the task, (as the one mentioned above did) but notes related to how much they were pulling the spring of the gun, what direction was used, what happens if direction is fixed or speed is changed, are missing. Results suggest that making notes related to the structure of the task is very rare, as in the case of Balancing.

Commonalities :

The number of children who do not make any notes is noteworthy; particularly in the more difficult tasks, Balancing and Rolling (see Chapter 6). The numbers rise when, as in Black Boxes and Floating, they make notes to help their memories (answers categorized as 'not relevant' for solving the task). Children who make relevant notes going beyond the intention to help memory, are very few. The APU (1988) have found that children's recording of findings is frequently 'disorganised and descriptive', and that 'appear to to feel little obligation to record the results ...'; something similar to this study.

It is interesting to note, however, that there are signs of this 'process' being consistent across tasks (see Chapter 7).

Levels of performance :

To make relevant notes ('good') for all the tasks seems quite uncommon; being more common when notes are to help memory ('partially good'), as for Black Boxes and Floating (see Figure 8.6.2). It is interesting that Balancing and Rolling, which have the most complex search spaces, have the largest number of children making no notes at all (see Section 8.1).

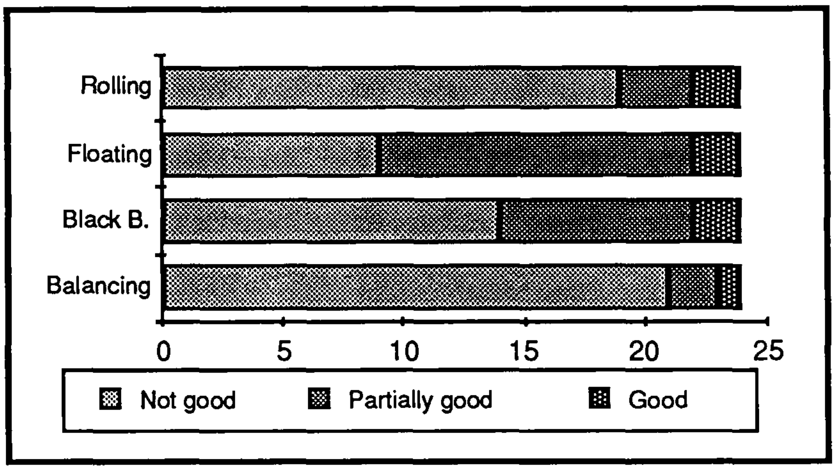


Figure 8.6.2: levels of performance on making notes in all the tasks.

Chapter 9.
DO PROCESSES RELATE TO PERFORMANCE IN SOLVING THE TASKS?

In previous chapters questions of the existence of processes across tasks, and of the cognitive demands within tasks have been addressed. This chapter discusses relationships between children's performance on processes (in both versions of the tasks) to their successes in solving the problem in the Investigation plus Goal tasks. The purpose is to see whether certain processes have a particularly close relationship to success in making an investigation, which might suggest that those processes could be considered as relevant to success on the task. If so, this could give greater predictive validity to the 'processes' elicited in the Structured tasks.

The analysis will follow the model of Chapter 7; using 2x2 contingency tables and $\ln \alpha$ as an indication of the sign of the relationship. Scores already explained in Chapter 5 and used in the subsequent chapters will be used.

9.1 DO SPONTANEOUS PROCESSES RELATE TO SUCCESS IN THE INVESTIGATION?

The question here is whether processes seen in the Investigation plus Goal tasks relate to performance in solving the problem posed by these tasks. These processes were 'spontaneous', in the sense that they were not responses to specified questions.

The 'processes' involved in all tasks are Searching Space (SS), Controlling Variables (CV), Replicating (R) and Making notes (Mn); except for Black Boxes that has SS1 (searching within boxes), SS2 (searching across boxes). Their relationships with success can be seen in Table 9.1.

Searching space:

In Black Boxes, all children tried all boxes, but no children tried the alternative strategy (comparing boxes by making the same test), so no relationship can be seen. In all the other three tasks the sign of the relationship is positive, though the evidence for a relationship existing is not very strong (Floating is the strongest but still weak).

			SS		CV		R		Mn	
			0	1	0	1	0	1	0	1
B	0	11	6	5	---	11	11	---	9	2
	1	13	5	8	---	13	13	---	12	1
			11	13	---	24	24	---	21	3
F	0	13	10	3	1	12	12	1	12	1
	1	11	6	5	0	11	11	0	10	1
			16	8	1	23	23	1	22	2
R	0	9	5	4	---	9	6	3	6	3
	1	15	6	9	---	15	8	7	13	2
			11	13	---	24	14	10	19	5
			SS1		SS2					
BB	0	8	---	8	8	---	3	5	6	2
	1	16	---	16	16	---	12	4	8	8
			---	24	24	---	15	9	14	10

Code: SS = Searching space; CV = Controlling variables; R = Replicating; Mn = Making notes.

Table 9.1: contingency tables showing the relationships between 'spontaneous' processes derived from the Investigation plus Goal tasks and performance in solving the problem on such tasks.

Controlling variables:

No relationship can be seen in any task because of the unbalanced marginal totals (essentially all children succeed).

Replication:

The only tasks where a relationship could have appeared are Rolling and Black Boxes. The first has a hint of a positive relationship; the second of a negative one. The problem with 'replicating' in Black Boxes is that we do not know whether children repeat a test because they want to make sure of its outcome or because they want to know where the ball is (if they lost track of it).

Making notes:

Again, a relationship could only appear in Rolling and Black Boxes. The first is negative, and the second positive. Since Rolling is mainly a manipulative task, it may not be surprising that making notes is either irrelevant to, or negatively relatively related to success.

Analysis:

All four processes analysed here show little evidence that they could be taken as predictors of success in solving the problems posed by the investigation tasks. There is one case ('controlling variables') where no relationships can be found because of the unbalanced marginals, and two cases ('replicating' and 'making notes) that are patchy and inconsistent. The last case ('searching space') is the best: all its relationships are positive when a relationship can appear, although they are weak. Thus, there are no strong candidates among the 'spontaneous' processes to be considered as predictors of success.

9.2 DO REQUESTED PROCESSES RELATE TO SUCCESS IN THE INVESTIGATION TASKS?

The processes built in to the Structured tasks and the framework behind them, were intended to be relevant to success in the tasks. The question is, do they actually relate?

Processes will be looked at in groups: a) noticing and identifying variables, b) making generalizations, c) what-if reasoning and imagining causes and, d) understanding. The reason for grouping them in this way, is that each group focuses on similar aspects: noticing and identifying variables both deal with detecting either changes or factors (that in the end produce changes if manipulated); what-if reasoning and imagining causes deal with knowing how something happens (prediction) and why such things happen (causes). 'Making generalizations' and 'understanding' are not grouped with the other processes.

Noticing (NT, N12, N3) and identifying variables (VT):

Where relationships can appear, several are positive in sign, being strongest for Floating (see Table 9.2). They are negligible for Balancing and Black Boxes, in which identifying factors or changes is not enough to solve the problem. In Balancing, noticing and identifying variables also require an understanding of how weight and distance are related. In Black Boxes the arbitrary nature of identifying variables and the strategy favoured by children (where discrimination of distances, essential in comparing boxes, is not used) seem responsible for the lack of relation to success. Rolling presents three positive relationships but two are weak (noticing, affected by expectations is the strongest), maybe reflecting that noticing is not enough to succeed. Noticing different floating levels and the factor(s) responsible for them, seem to be related to succeeding in the task, although the relationships are not that strong.

			NT				VT	
B	0	11	5	6			1	10
	1	13	6	7			1	12
			11	13			2	22
			0	1			0	1
BB	0	14	6	8			14	---
	2	10	4	6			10	---
			10	14			24	---
			0	2			0	2
			N12		N3			
F	0	13	8	5	13	---	9	4
	2	11	4	7	11	---	3	8
			12	12	24	---	12	12
			0	2	0	2	0	2
R	0	9	6	3	2	7	5	4
	1	15	6	9	0	15	6	9
			12	12	2	22	11	13
			0	1	0	1	0	1

Code: T indicates that all subtasks were combined; N = Noticing; N12 = Noticing, combining subtasks 1 & 2; N3 = Noticing, subtask 3; V = Identifying variables.

Table 9.2: contingency tables showing the relationships between 'noticing' and 'identifying variables' with success in solving the problem in the Investigation plus Goal tasks.

Making generalizations (GT):

One of the relationships is negligible (Black Boxes), the other is negligible (Floating), while the other two have a positive sign (Balancing and

Rolling), but are weak (see Table 9.3). Making generalizations fails in relating to success if children do not know how to adjust floating levels accurately (as they did), and it is also of little use if they do not compare boxes as it has been argued (see Chapter 8). Making generalizations seems more important for Balancing (the strongest of all four relationships), because doing so with different distances is the clue in solving the task. For Rolling, although it might help to make generalizations (paths are curved, ball always goes down), maybe expectations and manipulative skills are more important.

			GT	
B	0	11	7	4
	1	13	5	8
			12	12
			0	1

			GT	
R	0	9	4	5
	1	15	5	10
			9	15
			0	1

F	0	13	12	1
	2	11	11	0
			23	1
			0	2

BB	0	14	9	5
	2	10	6	4
			15	9
			0	2

Code: T indicates that all subtasks were combined; G = Making generalizations.

Table 9.3: contingency tables showing the relationships between 'making generalizations' and success in solving the problem in the Investigation plus Goal tasks.

What-if reasoning (WRT, HRT) and imagining causes (CT, C13, C2):

Again, where there can be a relationship, its sign is generally positive (though not for Rolling) (see Table 9.4). With this exception, there seems to be some tendency for these processes, taken together as broadly similar, to have some relation to success. Being able to predict an outcome (basically to know what would happen if certain conditions are met) and being able to imagine a cause (basically an explanation of why something happens) seem important in achieving success.

			WRT		CT	
B	0	11	9	2	9	2
	1	13	2	11	8	5
			11	13	17	7
			0	1	0	1

			WRT		CT	
R	0	9	7	2	8	1
	1	15	12	3	9	6
			19	5	17	7
			0	1	0	1

			HRT		C	
BB	0	14	7	7	14	---
	2	10	4	6	10	---
			11	13	24	---
			0	2	0	2

			WRT		C13		C2	
F	0	13	8	5	9	4	13	---
	2	11	2	9	7	4	11	---
			10	14	16	8	24	---
			0	2	0	2	0	2

Code: T indicates that all subtasks were combined; WR = What-if reasoning; C = Imagining causes; HR = Hypothetical reasoning; C13 = Imagining causes, combining subtasks 1 & 3; C2 = Imagining causes, subtask 2.

Table 9.4: contingency tables showing the relationships between 'what-if reasoning' and 'imagining causes' with success in solving the problem in the Investigation plus Goal tasks.

Understanding (U):

It has been mentioned before in Chapter 7 that the nature of 'understanding' is rather varied, and in this sense not useful as a predictor. Nevertheless, to make the exercise complete, the results for this 'process' are given (see Table 9.5). There is a negative relationship (Balancing) and two strong relationships (Floating and Rolling). The first is not surprising, given the content of the question: if children see bricks as a continuous or discrete variable or simply as different objects, nothing really to do with predicting correctly the outcome. In the case of Floating, it shows the same strength as the relationship between 'identifying variables' and success. They have the same content, but differ in the way questions were formulated (one gives alternatives to select while the other is open). Lastly, Rolling shows a strong relationship with 'understanding', that in this case means the expectations children have about the trajectory; something relevant to the solution of the problem.

			U	
B	0	11	6	5
	1	13	9	4
			15	9
			0	1

			U	
F	0	13	9	4
	2	11	3	8
			12	12
			0	2

			UT	
R	0	9	8	1
	1	15	7	8
			15	9
			0	1

Code: T indicates that all subtasks were combined; U = Understanding.

Table 9.5: contingency tables showing the relationships between 'understanding' and success in solving the Investigation plus Goal tasks.

Analysis:

The inspection of the relationships reveal that some processes do not relate consistently with success in the tasks: 'noticing-identifying-variables' and 'making generalizations'. They show a mixture of negligible, moderate and one or two strong relationships, casting doubts on their power as predictors. The case of 'understanding' is unique in the sense that, although it differs in nature for the three tasks, it suggests in one case that children's expectations are important for the solution of this task. This leaves 'what-if-reasoning-imagining causes' as the most viable candidate for predictor of success in the tasks, given the consistency of the positive relationships - when they can appear - and the nature of the processes involved.

Summary:

These results tend to suggest that:

i) 'natural processes', those derived from the Investigation plus Goal tasks, seem not to relate to performance in solving the problem; except probably for 'searching' the space of the task.

ii) requested 'processes', those built in the Structured tasks, seem to show more positive relationships with success in solving the problem and, significantly, the most consistent pattern is shown by those processes which tell us about how things happen and why.

These results seem to give some validity, though only in some cases, to the 'processes' coming from the Structured tasks and to the framework behind them.

Chapter 10.
CONCLUSIONS.

Emphasis on each of the elements of the triad: tasks, 'processes' and children's behaviour (see Figure 1.2), led to the analysis of data in the three chapters 6, 7 and 8 (Chapter 9 is the relation between 'processes' and the investigation tasks). Chapter 6 deals with the main pedagogical features of the tasks, Chapter 7 brings out the main problems in defining and eliciting 'scientific processes' in the tasks, and Chapter 8 describes children's performance on the tasks in the light of these 'processes' (see Figure 1.2).

This summary of the findings will follow the same structure in relation to the tasks, findings will be in two groups: seeing tasks a) as potentially pedagogic activities and b) as potentially diagnostic devices. The discussion of 'processes' will concern: a) problems of defining and b) problems of eliciting them. Findings in relation to children's behaviour will look at them as reflecting aspects of how children think and act in practice.

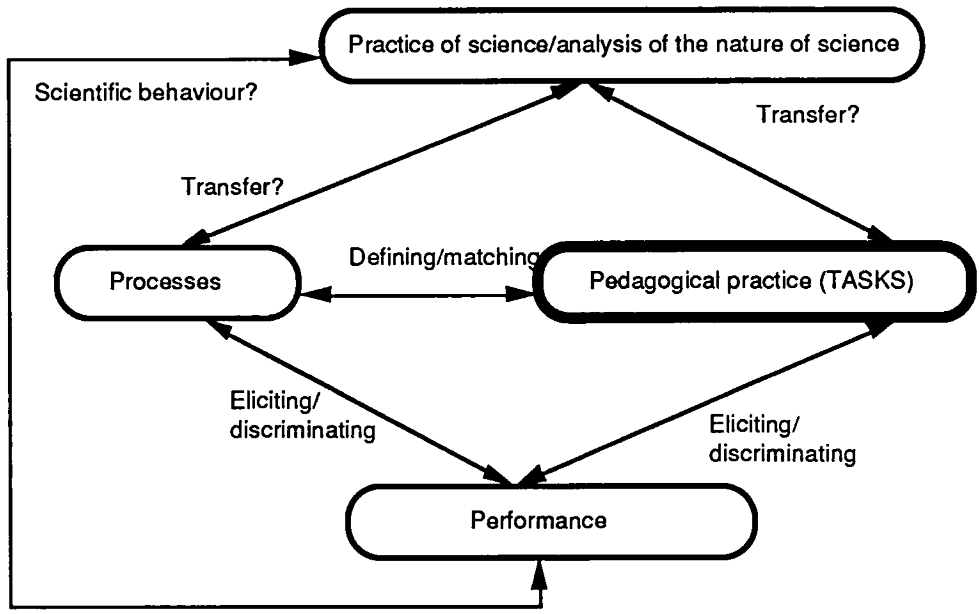


Figure 10.1: critical model of defining and eliciting children's 'scientific' behaviour.

Figure 10.1 is the same as Figure 1.2 of Chapter 1, where it served to account for the conception of the research. The following short explanation of this figure may help to clarify the logic behind the model and the sequence of exposition proposed above for the conclusions.

The main idea is that each step in eliciting children's 'scientific' behaviour on some predefined 'scientific' 'processes' should be questioned;

that none should be taken for granted. By contrast, the (perhaps extreme) hypothetical model for eliciting such processes in Figure 1.1 of Chapter 1, takes for granted: a) that processes deriving from philosophy of science belong to both scientific and educational practice, b) that such processes exist across domains, with their existence dependent only on the instrument's power of discrimination and, c) that children's performance is describable in the same qualitative terms as that of scientists. The intention has been to put all these assumptions into question.

In relation to the first assumption, it has already been argued in Chapter 1 that educational and scientific practices are different. As for the second, the problems of defining and eliciting processes across several domains will be shown. In connection with the third, it will be argued that children's reasoning should be described in terms of what it comes from - commonsense reasoning - not in terms of what for some it may in the future become - scientific reasoning.

This appears to ignore the findings on the tasks. In fact the discussion below will start with them, giving emphasis to the pedagogic significance of the whole exercise of 'eliciting scientific processes'.

10.1 TASKS.

Findings under this heading are of basic pedagogic importance in stressing issues concerning eliciting 'processes' starting from feasible and practical activities, based on and in accordance with teacher's experience and knowledge. That is, they concern tasks as the pedagogical medium for realizing 'processes' as educational objectives.

10.1.1 TASKS AS POTENTIALLY PEDAGOGIC DEVICES.

The Investigation plus Goal tasks were designed to be appropriate as classroom activities. Although not tested as such, teacher's opinions of them rated them quite favourably as: a) possible for children to do and as, b) worth doing (Section 3.3).

+ The tasks could be successfully performed by children, and show some power of discrimination (Section 6.3). This is of importance if the tasks, especially in the Investigation plus Goal ones, are to be used as teaching devices. It is important for the Structured ones for diagnostic reasons. Additionally, levels of success between both versions of the tasks are fairly consistent (section 6.6), suggesting that a link between diagnostic and teaching instruments might be possible; something to be welcomed in the attempt to bring assessment and teaching nearer to one another.

+ Teacher's expectations and rating of the scientific processes studied have some tendency to coincide with what children can actually do and give some support to them as educational objectives, though in different degrees (Section 6.8).

+ The investigation plus Goal tasks permitted the observation of the 'process' making notes; and the extent and systematism of children's searches of a space of variables (Section 5.1). The latter, gives information about children's spontaneous ways of organizing the task for themselves (Section 8.1) which is impossible to achieve with the Structured tasks alone. The Structured tasks, however, were needed to allow some processes to be investigated, such as 'making generalizations' and 'what-if' reasoning. The analysis suggested that other 'processes' like 'noticing' are more content bound (Section 7.1). This conclusion, however, remains only weakly supported because of the necessary restricted nature of the study, although encouraging enough to pursue the matter further.

+ It was possible to construct reasonably matching tasks for the Investigation plus Goal approach, even in the case of Black Boxes which lacks a clear scientific content. The structures of the search spaces could not be made identical but some were similar (Section 8.1).

+ It appears that it may be possible to control the difficulty of the Investigation plus Goal tasks, where the most difficult tasks had the more complex search spaces, by controlling the size of the space (Sections 6.4 and 5.1.1).

10.1.2 TASKS AS POTENTIALLY DIAGNOSTIC DEVICES.

+ Although the tasks were not developed as a battery of tests, children were reasonably consistent in their level of performance within versions across different contents (particularly in the Structured ones) and between versions (same content); suggesting that children who perform well in one task, do the similarly in other(s), giving an idea of what children can and cannot do (Sections 6.5 and 6.6). Evidence from more tasks and a larger sample would be needed to adequately secure this conclusion.

+ The Structured tasks seem to show some evidence of the importance of cognitive 'processes' like 'making generalizations', 'what-if' reasoning and 'identifying variables'. Further study would be needed to see whether these processes have a similar importance in other tasks, and whether they maintain a similar correlation with extent of search. Identifying a variable seen as a causal agent affecting another factor, making generalizations as the way to establish the functional relationship between factors, and 'what-if' reasoning as the possibility of predicting the outcome of such relationship, look to be as an encouraging start in constructing a framework for processes. Some tasks, like Rolling, require a perhaps excessive amount of manual dexterity (presence of so many attempts after failure within the same configuration) as well as relying heavily on children's expectations (Chapter 7).

+ It proved difficult, but not impossible, to match the Structured tasks in order to elicit similar 'processes' for different contents. The best match is between Floating and Balancing, where more genuine variables (physical quantities) were represented (Chapter 7), although they differ in the complexity of the search space (Section 8.1). Even in the case of a supposedly content-free task like Black Boxes, matching was partially possible: e.g. 'making generalizations' and 'what-if' reasoning (Chapter 7).

+ Processes, particularly those from the Structured tasks, appear to have some correlation (though not strong) with performance in solving the problem posed in the Investigation plus Goal tasks, suggesting that they might be taken as diagnostic elements; with 'what-if-reasoning-imagining causes as the best candidate (Chapter 9). But this conclusion would require a more clear cut experimental design than was possible here, to be established with any confidence.

10.2 PROCESSES.

Defining processes, that is representing some desired pedagogical objectives as 'scientific' 'processes', presents problems. It cannot be taken for granted that they can automatically or easily be implemented across tasks. Eliciting children's 'scientific' performance faces problems such as whether a process intended to be the same for different tasks does in fact have the same nature when performed in these different contexts.

10.2.1 PROBLEMS OF DEFINING.

+ It is an empirical question to define 'scientific processes', by trying to combine theoretical as well as pedagogical criteria. Problems were found when trying to define the same process for different tasks. For example 'what-if' reasoning for Rolling, turned out to be difficult to realize in practice because children's notion of direction was very limited. This is an example of practical problem of defining a process (Chapter 5).

+ Processes may have different sources, as in the case of 'making notes'; it would hardly appear on a list of processes derived from philosophy of science, but nevertheless arguably has some importance. The need for some 'external memory' when dealing with complex tasks is difficult to doubt.

+ The problem of the compromise between some attractive processes deriving from philosophy of science and their feasible representation in a task as pedagogical objectives, will become clearer when findings on performance are discussed. But it can be said in advance that processes derived from philosophy of science do not have an automatic existence in educational practice, and because both practices (science as a professional activity and education) are different, a 'didactic transposition' needs to be made (Chapter 2).

+ The notion of 'searching a space of variables' certainly presented some problems in adapting it to different tasks, but it was nevertheless found to give potentially valuable information related to the extent of the search during the investigation task (Section 5.1.1).

10.2.2 PROBLEMS OF ELICITING.

+ The difficulty of matching processes across tasks is particularly clear in the case where a process like 'identifying variables', intended to be the same across tasks, turns out to have in reality very different connotations for different tasks: physical quantities (Balancing and Floating), manipulative dexterity (Rolling) and what to pay attention to or test (Black Boxes). Another example is 'searching': a process such as how completely children investigate possibilities can be defined by comparing what children try with what they need to try. However, tasks differ both in how complex is the space of combinations of possibilities and in how clearly the nature of the task directs attention to different possibilities. The task may make possibilities evident and easy to keep track of, or it may not. Thus a simple measure of the ratio of possibilities tried to those needed may well be less useful than might appear.

+ It is an empirical matter to find processes which have similar discriminating power, in order even to be able to look for consistency across tasks.

+ Some 'processes' appeared more content dependent than others. 'Noticing' seemed to be rather content dependent, in agreement with those who see observation as content laden. Others, like 'making generalizations' and 'what-if reasoning' appeared to be less content dependent; as if abstracting 'behaviours' of a phenomenon from a particular case to a wider range of cases and, predicting outcomes given some conditions, are less content dependent due to the necessity of working at some abstract level. Work with a larger sample would be needed to give further support to these tentative conclusions.

+ A correlational approach (based on variability, thus norm-referenced) presents a paradox when it is found empirically that processes may tend to be very coherent across contents ('imagining causes', Section 7.1.1.4) but with low or very low variability. The only substitute would be a criterion-referenced definition of a process. For this we would need to know what would be for a child to have or not to have such a process or ability and, this would perhaps lead us to a developmental study in the existence of processes in children.

10.3. CHILDREN'S BEHAVIOUR.

When the attempt is made to analyze children's performance in relation to the source of inspiration (philosophy of science) used to define 'scientific processes', it becomes clear that the nature of children's actual behaviour is often different from that of professional scientists; something already signalled from a theoretical perspective in Chapter 1. The attempt to characterize children's performance in terms of commonsense reasoning is intended to stress such qualitative differences.

10.3.1 SOME COMMONSENSE FEATURES OF CHILDREN'S BEHAVIOUR.

+ In organizing their search of possibilities (Floating and Black Boxes) children may reason in terms of parts rather than wholes; that is, they do not tend to see density in one beaker as a 'value' of a quality to be varied but as an object (beaker) for example.

+ When doing the investigation, children seem to tend to prefer searching simple arrangements (Balancing and Rolling), although they tend to avoid the obvious ones; in the case of Balancing they prefer symmetrical arrangements.

+ Children do not tend to replicate results to make sure of their findings.

+ In processes like 'making generalizations', children can make what can be taken as plausible generalizations, nevertheless they do not tend to make explicit the assumptions under which such generalizations are valid. This makes it less plausible to count such process as being 'scientific' in nature. Similar problems appear with 'identifying variables' when children identify as factors such things as 'bricks', 'stripes', 'shots', 'dirty water', 'pulling', or 'pointing the gun', which include, besides variables, objects and events or actions.

+ Identifying variables appears to have two salient features. Children tend to identify mostly, and more spontaneously, factors related to action (weight, velocity), rather than factors that can only be chosen (distance, density, direction).

Some suggestive conclusions may be possible. It appears that the tasks used here tend to elicit discrete 'variables' (bricks, shots) and qualitative ones (pulling hard the spring, pointing high or beakers), but not continuous variables. And when children are asked if the amount of plastic affects the balance, they seem to prefer discrete indicators (number of dots on top of the bricks) and qualitative ones (bricks of the same size).

+ Children seem to it find difficult to give justifications for their identification of what may count as a variable, and for their predictions.

+ Children seem to find it easier to give causal explanations in terms of something given by the senses or experiences (descriptions of past experience for example) rather than in terms of things which have less to do with the senses (something pushing down in the case of Balancing); contrasting with the more abstract and rational explanations given in science.

+ Children's performance seems sometimes to be markedly affected by their expectations; the best candidate for this may be 'noticing' a ball rolling along a path which cannot in fact occur.

10.4 SUMMARY.

These findings, grouped around tasks, 'processes' and performance seem to support the following propositions:

a) that feasible pedagogical activities acceptable to teachers can be used to elicit at least some 'scientific' processes, in such a way as to bring teaching and assessment nearer.

b) that the nature of scientific processes is not self evident in an educational context.

c) that it is common to find 'processes' being content dependent, often because of cognitive differences between tasks. Some 'processes' however may be less content dependent.

d) that a combination of correlational and developmental (criterion referenced) evidence may be needed to establish whether a 'process' can be seen as existing.

e) that scientists' and children's performance are qualitatively different.

REFERENCES.

ADEY, P. & HARLEN, W. (1986) "A piagetian analysis of process skill test items", *Jour. of Res. in Sc. Teach.,* Vol. 23, No. 8, 707-726.

ALBERTS, R. V. et al. (1986a) "The development of standardized tests for experimental work in schools in the Netherlands", *Eur. Jour. of Sc. Ed.*, Vol. 8, No.2, 135-143.

ALBERTS, R. V. et al. (1986b) "The assessment of practical work: A choice of options", *Eur. Jour. of Sc. Ed.*, Vol. 8, No. 4, 361-369.

AMERICAN ASSOCIATION FOR ADVANCEMENT OF SCIENCE (A.A.A.S.)(1967)*Science - A Process Approach, part six* (4th experim. ed.), Washington, D. C., AAAS Miscellaneous Publication 67-11.

AMERICAN ASSOCIATION FOR ADVANCEMENT OF SCIENCE (A.A.A.S.) (1968) *Science - A Process Approach. An Evaluation Model and its application.* (Second Report), AAAS Miscellaneous Publication 68-4.

AYLWARD, G. & EGGLESTON, J. (1986) *Developing skills in solving scientific problems through group practical tasks,* Jakarta.

BLACK, P. (1990) "APU Science - the past and the future", *School Science Review*, Vol. 72, No. 258, Sept., 13-28.

BLACK, P., HARLEN, W. & ORGEE, T. (1984) *Standards of performance expectations and reality. A study of the problem of interpreting the APU science surveys.* London: DES/APU (Occasional paper 3).

BRYCE, T. et al. (1983) *Techniques for the assessment of practical skills in foundation science,* London, Heinemann Educational Books.

BRYCE, T. & ROBERTSON, I. (1985) "What can they do? A review of practical assessment in science", *Studies in Science Education,* Vol. 12, 1-24.

CAREY, S. (1985) *Conceptual change in childhood*, Cambridge, Mass., MIT Press

CHEVELLARD, Y. (1985) *La transposition didactique du savoir savant au savoir enseigné,* La Pensée Sauvage, Editions Juin.

CHIPMAN, S. F. & SEGAL, J. W. (1985) "Higher Cognitive Goals for Education: an introduction". In Segal, J. W., Chipman, S. F. & Glaser, R, *Thinking and learning skills,* Vol. 1: Relating instruction to research, Hillsdale N. J., LEA.

CHIPMAN, S. F. & SEGAL, J. W. (1985) "Higher Cognitive Goals for Education: an introduction". In Segal, J. W., Chipman, S. F. & Glaser, R., *Thinking and learning skills*. Vol. 2: Research and Open Questions, Hillsdale N. J., LEA.

DEPARTMENT OF EDUCATION AND SCIENCE (D.E.S.)/A.P.U. (1981)*Science in schools. Age 11: Report No. 1*, London, H.M.S.O.

DEPARTMENT OF EDUCATION AND SCIENCE (D.E.S)/A.P.U. (1983)*Science in schools. Age 11: Report No. 2*, London, H.M.S.O.

DEPARTMENT OF EDUCATION AND SCIENCE (D.E.S)/A.P.U. (1984)*Science in schools. Age 11: Report No. 3*, London, H.M.S.O.

DEPARTMENT OF EDUCATION AND SCIENCE (D.E.S.) (1985) *Science 5-16: A statement of policy*, London, H.M.S.O.

DEPARTMENT OF EDUCATION AND SCIENCE (D.E.S.)/APU (1988)*Science at Age 11. A Review of APU Survey Findings 1980-84*, London, H.M.S.O.

DEPARTMENT OF EDUCATION AND SCIENCE (D.E.S.)/APU (1989)*National Assessment: The APU Science Approach*, London, H.M.S.O.

DONNELLY, J. & GOTT, R. (1985) "An assessment-led approach to process in the science curriculum", *Eur. Jour. of Sc. Ed.*, Vol. 7, No. 3, 237-251.

DORAN, L. (1978a) "Measuring the processes of science objectives", *Science Education*, 62 (1), 19-30.

DORAN, L. (1978b) "Assessing the Outcomes of Science Laboratory Skills",*Science Education.*, 62(3), 401-409.

DRIVER, R. (1983)*The pupil as scientist*, Milton Keynes, Open University Press.

EGGLESTON, J. (1984) *A summative evaluation of Pemantapan Kerja Guru*, Jakarta.

EVANS, J. (1982) *The Psychology of Deductive Reasoning*, London, Routledge & Kegan Paul

FAIRBROTHER, B. (1989) "Problems in the assessment of scientific skills". In Wellington, J. *Skills and Processes in Science Education. A Critical Analysis*, London, Routledge.

FINLEY, N. (1983) "Science processes", *Jour. of Res. in Sc. Teach.*, Vol. 20, No. 1, 47-54.

GAGNE, R. (1962) "The acquisition of knowledge", *Psychological review*, Vol. 69, No. 4, 355-365.

GAGNE, R. (1963) "The learning requirements for enquiry", *Jour. of Res. in Sc. Ed.,* Vol. 1, 144-153.

GAGNE, R. (1965)*The psychological bases of Science-A Process Approach,* Washington D.C., American Association for the Advancement of Science Miscellaneous Publication 65-8.

GARNETT, P. et al. (1985) "Reasoning abilities of secondary school students aged 13-16 and implications for the teaching science", *Eur. Jour. of Sc. Ed.,* Vol. 7, No. 4, 387-397.

GREEN, B. Mc CLOSKEY, M. & CARAMAZZA, A. (1985) "The relation of knowledge to problem-solving, with examples from kinematics". In Segal, J. W., Chipman S. F. & Glaser, R., *Thinking and learning skills,* Vol. 2: Research and Open Questions, Hillsdale N. J., LEA.

HARLEN, W. (1975) *Science 5-13: A Formative Evaluation,* London, Macmillan Education.

HELLINGMAN, C. (1982) "A Trial List of Objectives of Experimental Work in Science Education", *Eur. Jour. of Sc. Ed.*, 4(1), 29-43.

HODSON, R. (1986a) "Rethinking the role and status of observation in Science Education", *Journal of Curriculum Studies,* Vol. 18, No. 4, 381-396.

HODSON, R. (1986b) "The nature of scientific observation", *School Science Review,* 17-27.

JENKINS, E. W. (1989) "Process in science education: an historical perspective". In Wellington, J. (Ed.), *Skills and Processes in Science Education,* London, Routledge.

JOHNSON, S. (1987) "Assessment in Science and Technology", *Studies in Science Education,* 14, 83-108.

KARPLUS, R. & THIER, H. (1967) *A New Look at elementary School Science. Science Curriculum Improvement Study,* Chicago, Rand McNally & Company.

KNORR-CETINA, K. D. (1981)*The manufacture of knowledge. An essay on the constructivist and contextual nature of science,* Oxford, Pergamon Press.

KRUGLAK, H. (1954) 'The measurement of laboratory achievement'. *American Journal of Physics.* Vol. 23, 82-87.

KUHN, D. & BRANNOCK, J. (1977) "Development of the isolation of variables scheme in experimental and 'natural experiment contexts", *Developmental Psychology,* Vol. 13(1), 9-14.

LANG, H. (1982) "Criterion-Referenced tests in Science: an investigation of reliability, validity, and standards-setting", *Journal of Research in Science Teaching,* 19(8), 665-674.

LINDQUIST, E. F. (1953) *Design and analysis of experiments in Psychology and Education,* Boston: Houghton Mifflin.

MARTINAND, J. L. (1983) "Questions for research: Referring to a practice and possible types of scientific activities in schools", in *International summer workshop: Research on Physics Education,* La Londe les Maures, Editions du CNRS (Centre National de la Recherche Scientifique).

MARTINAND, J. L. (1986) *Connaître et transformer la matière. Des objectifs pour l'initiation aux sciences et techniques,* Berne, Peter Lang.

MC DERMOTT, L. (1983) "Critical review of research in the domain of mechanics", in *Research on Physics Education: proceedings of the first international workshop,* La Londe les Maures, Editions du CNRS (Centre National de la Recherche Scientifique).

MC LEOD, R. J. et al. (1975) 'The development of criterion-validated test items for four integrated science processes', *Journal of Research in Science Teaching,* Vol. 12, No. 40, 415-421.

MILLAR, R. & DRIVER, R. (1987) "Beyond processes", *Science Education,* 14, 33-62.

MOLITOR, L. & GEORGE, K. (1976) "Development of a test of science process skills", *Journal of Research in Science Teaching*, Vol. 13, No. 5, 405-412.

NEDELSKY, L. (1965) *Science teaching and testing.* Brace & World, Harcourt.

NUFFIELD JUNIOR SCIENCE PROJECT (1970) *Teacher's Guide 1,* London, Collins.

OGBORN, J. (1988a) *The nature of science and its implications for science for all.* In Sequeira, M., Leite, L. and Freitas M. (Eds.) *Actas de I Encontro Sobre Educação em Ciências,* Braga, May, 1987, Universidade do Minho, Portugal.

OGBORN, J. (1988b) *A map of science.* Paper invited by the National curriculum Science Working Party. Unpublished; personally communicated.

OGBORN, J. (1990) Personal communication.

PADILLA, M. et al. (1983) "The relationship between science processes skills and formal operational thinking abilities", *Jour. of Res. in Sc. Teach.,* 20(3), 239-246.

PHILLIPS, D. C. & KELLY, M. E. (1975) "Hierarchical theories of development in Education and Psychology", *Harvard Educational Review*, Vol. 45(3), 351-375.

POLLARD, P. (1982) "Human reasoning: some possible effects of availability",Cognition, 12, 65-96.

POSNER, G. J. & STRIKE, K. A. (1976) "A categorization scheme for principles of sequencing content, Review of Education Research, Vol. 46(4), 655-690.

RAVETZ, J. R. (1971) *Scientific knowledge and its social problems*, Oxford, Oxford University Press.

ROSS, J. & MAYNES, F. (1983) "Development of a test of experimental problem-solving skills", *Jour. of Res. in Sc. Teach.*, Vol. 20, No. 1, 63-75.

SCHOOLS COUNCIL (1972) *With Objectives in mind. Guide to Science 5-13*, London, Macdonald Education.

SCIENCE PROCESSES AND CONCEPT EXPLORATION PROJECT (SPACE) (1990) *Evaporation and Condensation. Research Report*, Liverpool, Liverpool University Press.

SCREEN, P. (1986) *Warwick process science*, Southampton, Ashford Press.

SHAYER, M. & ADEY, P. (1983) *Towards a science of science teaching*, London, Heinemann Educational Books.

SHAW, T. (1983) "The effect of a process-oriented science curriculum upon problem solving ability", *Science Education*, Vol. 67, No. 5, 615-623.

TOBIN, K. (1984) "Reasoning ability of upper primary science pupils, *The Australian Science Teachers Jour.*, Vol. 30, No. 1, 75-81.

WELLINGTON, J. (Ed.)(1989) *Skills and processes in science education. A introduction*, London, Routledge.

WOOLNOUGH, B. E. (1989) "Towards a holistic view of processes in science education (or the whole is greater than the sum of its parts and different)". In *Skills and processes in Science Education*, Wellington (Ed.) London, Routledge.

YEANY, R. et al. (1986) "Analyzing hierarchical relationships among modes of cognitive reasoning and integrated science processes skills", *Jour. Res. in Sc. Teach.*, Vol. 23, No. 4, 277- 291.

ZIMAN, J. (1984) *An introduction to science studies; the philosophical and social aspects of science and technology*, Cambridge, Cambridge University Press.

BIBLIOGRAPHY.

ADEY, P. (1987) "Science develops logical thinking- doesn't it? Part 1: Abstract thinking and school science", *Science School Review,* June.

BENNETT, N. et al. (1984) *The quality of pupil learning experiences,* London, Lawrence Erlbaum Associates.

BLACK, H. D. DOCKRELL, W. B. (1984) *Criterion-referenced assessment in the classroom,* Edinburgh, The Scottish Council for Research in Education.

BLENKIN, G. & KELLY, A. (1981) *The primary curriculum,* London, Harper & Row.

CARRE, A. & DRIVER, R. (leaders)(1983) "The evaluation of individual performances" (Workshop No. 1), in *Research on physics education: proceedings of the first international international workshop,* La Londe les Maures, Editions du CNRS (Centre National de la Recherche Scientifique).

ERICSSON, K. A. & SIMON, H. A., "Sources of evidence on Cognition: a historical review". In Merluzzi, T., Glass, C. & Genest, M. *Cognitive Assessment,* New York, Guilford Press.

ERICKSON B. H. & NOSANCHUK T. A. (1985) *Understanding Data. An introduction to exploratory and confirmatory data analysis for students in the social sciences,* Milton Keynes, Open University Press.

FERGUSON, G. A. (1985) *Statistical Analysis in Psychology and Education,* (5th ed.), Singapore, McGraw-Hill.

FIENBERG, S. E. (1978) *The Analysis of Cross-Classified Categorical Data,* Cambridge, Mass., The MIT Press.

HACKING, I. (1983) *Representing and intervening. Introductory topics in the philosophy of natural science,* Cambridge, Cambridge University Press.

HARLEN, W. et al. (1977) *Match and mismatch. Raising questions: Leaders guide,* Edinburgh, Oliver & Boyd for the Schools Council.

HARLEN, W. (1983) *Guides to assessment in education, Science,* London, Macmillan.

HARLEN, W. (1985) *Teaching and learning primary science,* London, Harper & Row.

HARRE, R. (1988) *The philosophy of science. An introductory survey,* Oxford, Oxford University Press.

HODSON, D. (1982) "Is there a scientific method?", *Education in Chemistry,* Vol. 19, No. 4, 112-116

HODSON, D. (1985) "Philosophy of science, science and science education", *Studies in Science Education,* 12, 25-57.

HODSON, D. (1988) "Towards a philosophically more valid Science Curriculum", *Science Education,* 72(1), 19-40.

KAMII, C. & De VRIES, R. (1978) *Physical knowledge in Preschool Education: implications of Piaget's theory,* Englewood Cliffs, N. J., Prentice Hall Inc..

KARMILOFF-SMITH, A. & INHELDER, B. (1974-75) "If you want to get ahead, get a theory", *Cognition.* 3 (3), 195-212.

KEMPA, R. (1986) *Assessment in science,* Cambridge, Cambridge University Press.

KLAHR, D. & DUNBAR, K. (1988) "Dual space search during scientific reasoning",*Cognitive Science,* Vol. 12, 1-48.

KUSLAN, L. & STONE, A. (1968) *Teaching children science: an inquiry approach,* California, Wadsworth.

MILLAR, R. (1987) "Towards a role for experiment in the science teaching laboratory", *Studies in Science Education,* 14, 109-118.

MILLER, G. et al. (1960) *Plans and the structure of behavior,* New York, Holt, Rinehart & Winston, Inc.

NEWELL, A. & SIMON, H. A. (1972) *Human Problem Solving,* Englewood Cliffs, N.J., Prentice Hall Inc..

NITKO, A. (1980) "Defining 'criterion-referenced test'". In Berk, R. *A guide to criterion referenced test construction,* Baltimore, John Hopkins University Press.

OSBORNE, R. & FREYBERG, P. (1985) *Learning in science. The implications of children's science,* London, Heinemann.

PIAGET, J. (1974) *Success and understanding,* London, Routledge & Kegan Paul.

RUBINSTEIN, R. A. et al. (1984) *Science as cognitive process. Towards an empirical philosophy of science,* Philadelphia, University of Pennsylvania Press.

WELLINGTON J. (Ed.) (1989) *Skills and processes in Science Education: A critical analysis,* London, Routledge.

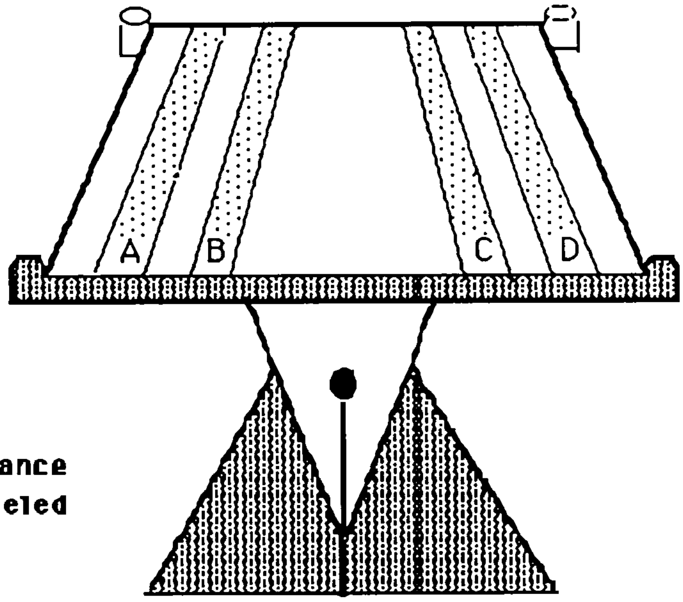
WHITE, R. (1988)*Learning science,* Oxford, Basil Blackwell.

APPENDIX A

INSTRUMENTS (Investigation plus Goal tasks).

BALANCING TASK

You have a **balance** and
a heap of **Lego bricks.**



Balance leveled

Suppose someone chooses some of your bricks, and puts them in one side of t balance (stripes "A" or "B").

Is there a way to choose bricks from the rest to put in the other side (stripes "C" "D"), so that the balance **is sure to be level**?

Can it be done in more than one way?

Try this for a while.

Then, We'll see how well you can do it.

I will pick some bricks and put them in one side of the balance.

You don't know which bricks I will pick and which stripe I will choose.

After that, you will have to choose the bricks and the stripe which you are sure \ balance my bricks. Then we will try, and see if you are right.

BALANCING TASK **Check list** SHAHEDA. 14min. ①

att	Set / Change conditions					Result		Vary				
	A	B	1*16	1*8	1*4	success	failure	C	D	1*16	1*8	1*4
	/		1_A		2_A	/			/		3_D	
	/	/		1_B 3_A	2_B		/	/	/	1_D		4_C
ch					1		/					
							/					3
	/				2_A				/		1_D	
+		/		1_B	1_A (1_B)	/		/			1_C	1_D 1_C
+				1_B 2_B	1 out 4_A		/			1_D 1_D	1_C 1_D	
+ (Rearrangement were weight were needed)							/					
	/	/	1_A	1_A	3_B		/	/	/		1_C 2_D	2_C
	/		1_A	1_A	3_A	/			/		3_D	3_D
	/		1_A				/	/			1_C	2_C
+						/					1_D	

BALANCING TASK **Check list** SHAHEDA.

att	Set / Change conditions					Result		Vary				
	A	B	1*16	1*8	1*4	success	failure	C	D	1*16	1*8	1*4
	/			4 A	1A	/			/	1D		5D

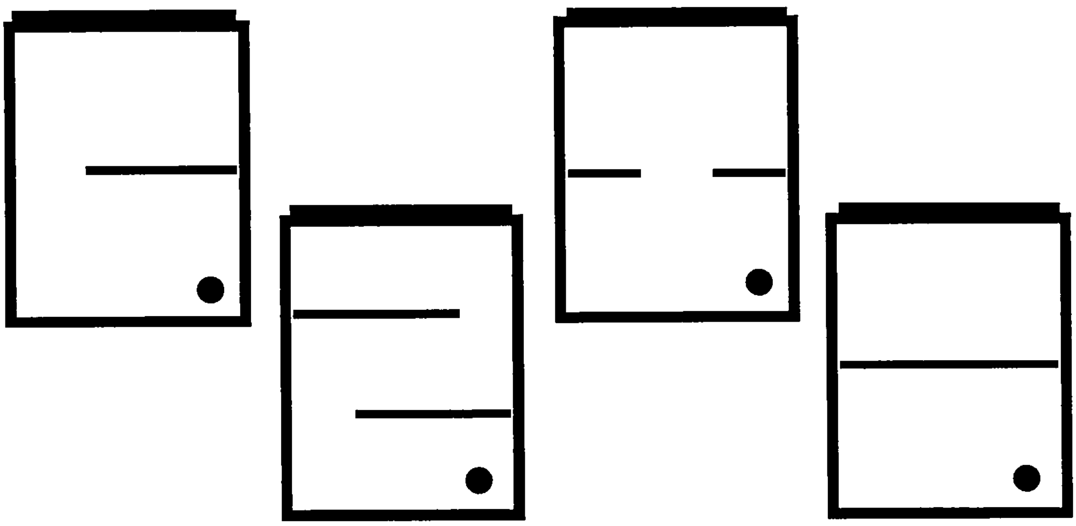
PROBL.
Rods on B
use A

PRED,
1×8
1×4

You have four **closed** boxes, each one has a marble inside.
They are coloured green, black, white and orange.

Don't open them.

All the boxes have different patterns inside.
Here are the pictures of all boxes.



TRY FOR A WHILE TO FIND OUT WHICH BOX IS WHICH.

Do it by **carefully** listening to the marbles move around the edges as you mo\ the boxes.

Then you should be able to know which pattern is inside each box.
You will only find out the patterns by making very **gentle movements** with tł boxes.

Tell me when you have finished moving the marbles inside the boxes and yc know the pattern inside each box.
After this, I will choose one of them and ask you to tell me what pattern inside it. You don't know which one I will choose. Then we will open tł box and see how good was your prediction.

BLACK BOXES TASK Check list RYAN 10 min.

①

att	Set/Change conditions (Struct)						Vary		
	Green box	Black box	White box	Orange box	top	bottom	left	right	obstacles
				✓				✓	
							✓		
									✓
							✓		
									✓
							✓		
							✓		
						–		✓	
							✓		
							✓		
							✓		
							✓		

(2)

BLACK BOXES TASK Check list RYAN

att	Set/Change conditions						Vary		
	Green box	Black box	White box	Orange box	top	bottom	left	right	obstacles
									✓
							✓		
									✓
							✓		
								✓	
							✓		
									✓
						✓		✓	
									✓
							✓		
	✓						✓		
							✓		

(3)

BLACK BOXES TASK   Check list   RYAN

a t t	Set/Change conditions						Vary		
	Green box	Black box	White box	Orange box	top	bottom	left	right	obstacles
								✓	
							✓		
									✓
								✓	
									✓
							✓		
							✓		
						—	✓		
							✓		
							✓		
								—	
							✓		

(4)

BLACK BOXES TASK   Check list   RYAN.

att	Set/Change conditions						Vary		
	Green box	Black box	White box	Orange box	top	bottom	left	right	obstacles
								✓	
		✓					✓		
								✓	
									✓
									✓
									✓
								✓	
						✓	✓		
									✓
									✓
									✓
									✓

(5)

BLACK BOXES TASK Check list RYAN

att	Set/Change conditions						Vary		
	Green box	Black box	White box	Orange box	top	bottom	left	right	obstacles
									✓
									✓
								✓	
							✓		
									✓
									✓
									✓
			✓			—		✓	
							✓		
									✓

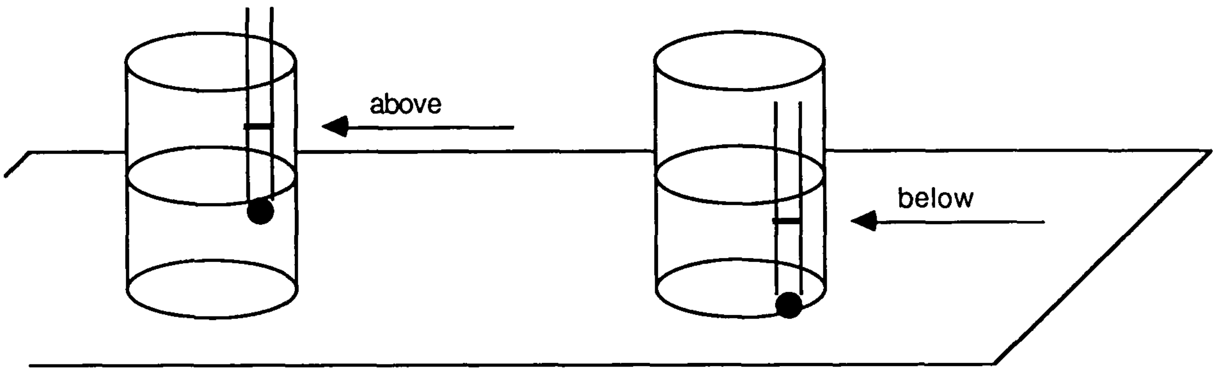
PROBL.

PRED. (Yes)

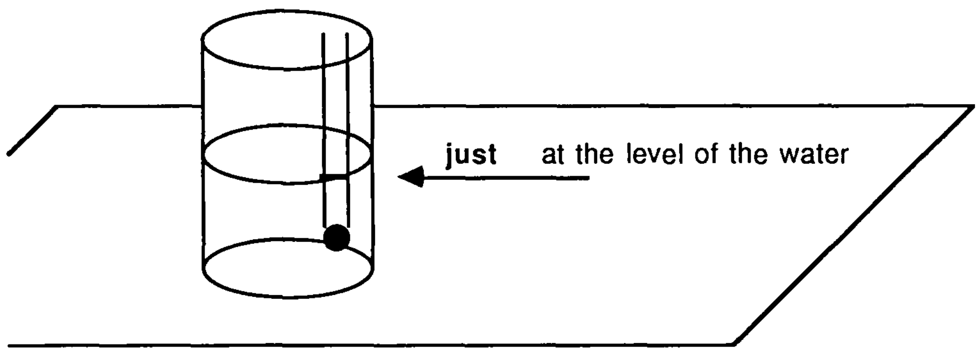
FLOATING TASK

Ships have painted on their hull a floating line. Its name is the Plimsoll line. Its sho the level at which the ship floats.

I have made two straws. One always floats with its line **above** the water, in all t beakers.
The other always floats **below**. Look and see how much steel shot I have put in then



You have a straw of your own.
Find out how much steel shot to put in, to make it float with the line *just* at the water.
Find out for all the beakers.



See how the straws float or sink for a while.

There is a straw, and steel shots for you to try.

Then, I am going to give you an empty straw with its Plimsoll line already drawn a ask you to **get the line *just* at the level of the water.**
You don't know which beaker I will choose.
Before you try, you will tell me how many steel shots do you need to get the Plims line ***just*** at the level of the water.Then, we will see how good is your prediction!

FLOATING TASK **Check list** BONNIE 14 min. No check { . above . below

att	Set/ Change conditions					Result		Vary	
	beaker A	beaker B	beaker C	beaker D	beaker E	success	fail	add shots	out shots
	✓							✓	
	✓						✓		✓ all
								✓	
							✓ near		✓
						✓			✓ all counting
								✓	
		✓					✓ very near		
			✓				✓ near		
				✓			✓	✓	
							✓	✓	
							✓	✓	
							✓ near		
					✓		✓	✓	
							✓	✓	
							✓	✓	

(2)

FLOATING TASK Check list BONNIE

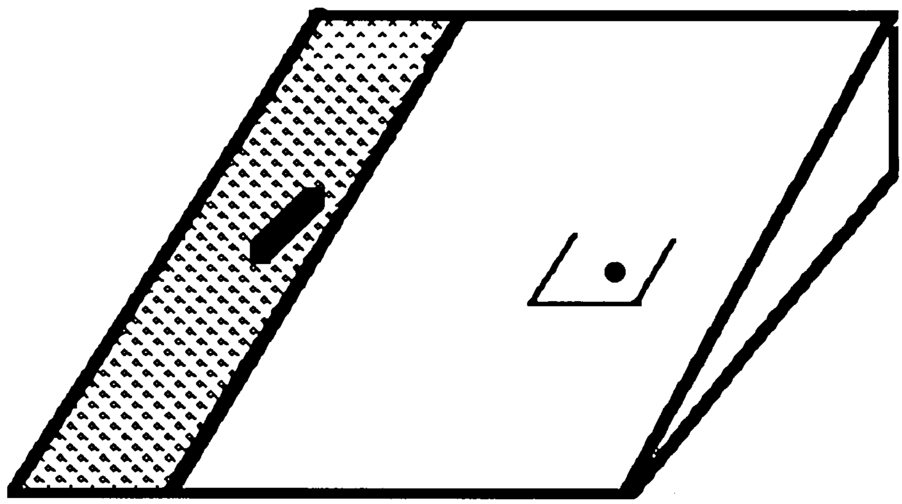
att	Set/ Change conditions					Result		Vary	
	beaker A	beaker B	beaker C	beaker D	beaker E	success	fail	add shots	out shots
							/ very near		

PROBL. How many in E.

PRED (14)

You have a 'gun', a ball, a trap, a tilted board, and a stripe at the left of the board.

Use the 'gun' to set the ball rolling, but place it on the stripe. You are not allowed to place the 'gun' outside the stripe.
The trap can be placed wherever you want on the board (not on the stripe).



Try for a while to get the ball inside of the trap.

Try it in as many ways as you can, but you are not allowed to touch the frame with the ball.

After that, I am going to set the 'gun' and the trap on the board, following the same rules you have followed, and to ask you to get the ball into the trap.

But, remember, you don't know which places I will choose for the 'gun' and trap. Therefore, try very hard in as many ways as you can in getting the ball into the trap.

At the end, before actually trying to get the ball inside of the trap, I will ask you to draw the path the ball would follow. Then we'll see how good your prediction was.

(1)

ROLLER BALL TASK RICHARD 17 min. Check list

att	Set / Change conditions: gun ↑	gun ↓	trap ↑	trap ↓	trap ←	trap →	trap ◇	trap ⊔	Result: success	fail	Vary: direction	velocity	none
		✓		✓		✓	✓			✓		✓	
										✓			
		✓	✓		✓		✓		✓				
		✓	✓			✓	✓			✓			
		✓		✓		✓	✓			✓		✓	
										✓			
		✓	✓			✓	✓			✓	✓		
									✓				
		✓		✓		✓	✓			✓		✓	
										✓			
		✓		✓	✓			✓		✓		✓	
										✓		✓	
										✓	✓	✓	
									✓				
		✓		✓	✓		✓		✓				
		✓		✓		✓		✓		✓		✓	

ROLLER BALL TASK RICHARD Check list

(2)

att	Set/ Change conditions: gun ↑	gun ↓	trap ↑	trap ↓	trap ←	trap →	trap ◇	trap ⊔	Result: success	fail	Vary: direction	velocity	none
										✓			
		✓	✓			✓	✓			✓	✓		
									✓				
										✓			
		✓	✓			✓	✓		✓				
									✓				
		✓	✓			✓	✓		✓			✓	
										✓			
		✓		✓		✓	✓			✓			
		✓	✓			✓	✓			✓	✓		
										✓			
		✓	✓			✓		✓		✓	✓	✓	
										✓			
		✓	✓			✓	✓			✓		✓	
										✓			
		✓	✓			✓	✓			✓		✓	

ROLLER BALL TASK   RICHARD   Check list

att	Set/Change conditions: gun ↑	gun ↓	trap ↑	trap ↓	trap ←	trap →	trap ◇	trap ⊔	Result: success	fail	Vary: direction	velocity	none
										✓	✓	✓	
										✓			
		✓		✓		✓	✓		✓				
									✓				
		✓	✓			✓		✓		✓		✓	
										✓	✓		
										✓		✓	
										✓			
		✓	✓			✓	✓			✓	✓		
										✓			
										✓	✓		
										✓	✓		
									✓				

PEREL.

He knows … not … from pulled … PEED

APPENDIX B

INSTRUCTIONS (Structured tasks).

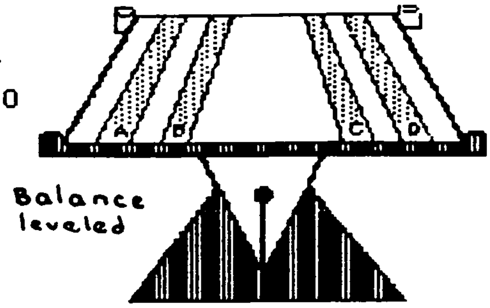
GAVIN (A)

53 min.

Section 1:

You have a set of bricks on different weights and colours, and a balance to compare them.

The picture shows how is the balance when has no bricks on it or have some with the same weight.



Put the yellow brick on stripe **A,** all the blues on stripe **D** , and the black on stripe **C.**

Tell me, what did you notice?

Inotice that when I put the yellow brick on ~~a~~ stripe A and all the blues on stripe D it ~~blnce~~ was lever when I put the black brick it tilt to the right

Put the yellow brick on stripe **A** , all the blues on stripe **C** , and the black on stripe **D.**

Tell me, what did you notice?

I notice when you change the ~~blu~~ place of the blues and the black it [illegible] up.

GAVIN (A)

Section 2:

Suppose your are trying to get the balance level, by using stripes **A** and **D**. You can do it by using different combinations of bricks.

What do you think makes the balance level?

A That the bricks have the same size.	**B** That the bricks have the same amount of dots on the top.
C That the bricks have the same amount of plastic.	**D**

Put here anything else you think makes the balance level.

Tick the box/es you agree with:

A	B	C
✓		

Section 3:

Put all the reds on stripe **B**.

Without actually trying, think:

what would be the best bricks to use to get the balance level, if you must put the bricks on stripe **D** ?

Write here what bricks you would use.

The black

Why do you think **that** is a good combination of bricks to get balance level?

brick balance diffend way if you put it diffend way.

balance

Diff way = diff distances.

GAVIN (2)

Put all the reds on stripe **A.**

Without actually trying, think:

What would be a good stripe to place the **yellow** and the two **blue** bricks to get the balance level?

Write here which stripe would you use.

C

Why do you think **that** is a good stripe to get the balance level?

because if you put the blue's on the D it wall balance tother but if you put the blue and the yellow on the D it wall slinght down to the right so I wall put the blue and yellow on C

Section 4:

You have tried to get the balance or scale level in different ways.

What is one thing that getting the scale to balance depends on?

The amounts of brick you put in each side.

amounts

What difference does it make?

if you put the equal amounts of brick on each side it wall balance tother but if you put more on one side and less on the other it wall tilt.

GAVIN (A)

What is **another** thing that getting the scale to balance depends on?

where you put ~~the~~ bricks on the scale; I mean the stripe you put the brick

↳ What difference does it make?

if you put the yellow brick on a and red's and blue's brick on stripc it woll balance the same but if you use D for the red's and bloe's it woll tilt to one side

Section 5:

You have noticed how different bricks in different places can make the balance level.

But tell me, is there something which is the **same** about all the ways to make the balance level?

To put the ~~same~~ same amounte of bricks on each side.

Is there anything **else** which is the same about all the ways to make the balance level?

if you put a brick on ~~the~~ strip a and you put the duber on strip c it wood balance.

GAVIN (A)

Section 6:

Tilt the balance carefully by pushing **A** with your finger.
Now, Put the **black** brick on stripe **A**.
Your finger and the brick have the same effect: they tilted the balance.
Tell me, why finger and brick had the same effect on the balance?

my finger put pressure on the left hen side of the sceall and so dose the brick

You can make the balance **level** by pushing both sides with your fingers.
Why does this do the same as having bricks which **do** balance?

because my finger put the same amounte of pressure on each side so dust the brick.

SALLY.

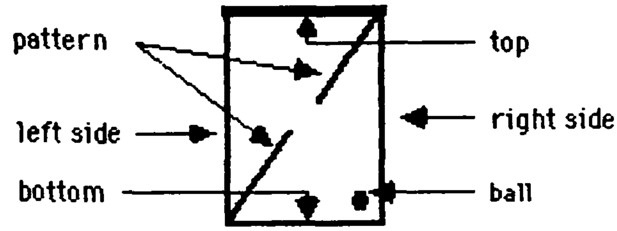
32min.

Is very difficult when they are thinking in terms of whole gap because they do not pay attention to the information given by obstacles.

Section 1:

You have 4 closed boxes that have different patterns inside and a ball for you to roll around. The boxes are like the one is in front of you. It is open for you to look.

The open box can be drawn like this if it is seen from above:



Hold the **green** box as I am going to give it to you.
The ball is on the **right top corner** of the box.
Move the ball along the right side by tilting the box up and down carefully several times.

Listen carefully at the ball moving.

Take now the **white** box and do the same as you did with the green one.

Tell me, what did you notice?

I notice that the ball gos faher in the green box then the write box

Hold the **green** box as I am going to give it to you.

The ball is on the **right bottom corner** of the box.

Move the ball along the right side tilting the box up and down carefully several times.

Listen carefully at the ball moving.

Take now the **orange** one and do the same as you did with the green one.

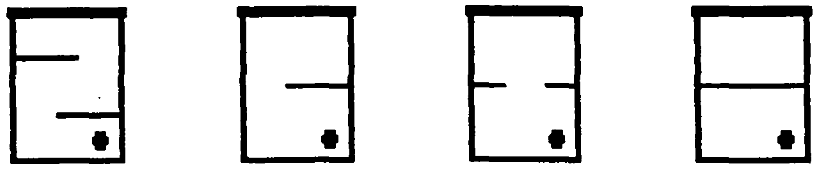
Tell me, what did you notice?

~~th~~ I notice that the ball didnt go so far then the write box

SALLY

Section 2:

The patterns you have in the 4 **closed** boxes are:



Suppose you have the problem of finding out which box is which by listening to how the ball rolls.

Which of these is the best way to do it?

1 To stick with one box until you discover which one it is. Then do the same with the other boxes.	**2** To choose one test (for example: running the ball down one side) and test this for all the boxes Then do the same with other tests.
3 To stick with one pattern until you can find which box has that pattern. Then do the same with all the patterns.	**4**

Tick the box you agree with:

Write how to do it if you don't agree with 1, 2 or 3, here.

1	✓	2		3	

Section 3:

Pick up the green box.

Try to get the ball into all 4 corners of the box.

Tell me when you have got the ball into each corner.

1. Right top corner	✓	3. Left top corner	
2. Left bottom corner	✓	4. Right bottom corner	✓

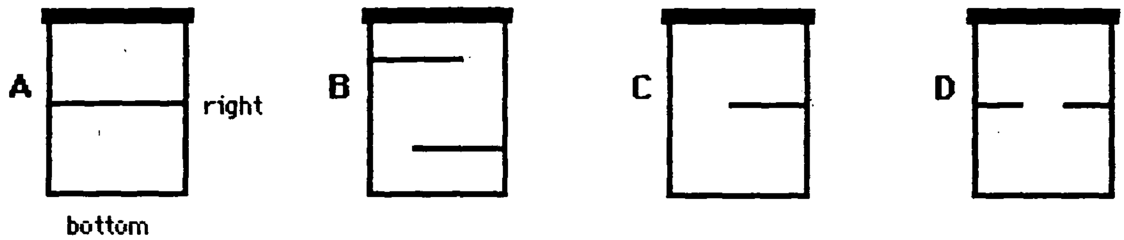
SALLY

Suppose you give the boxes problem of section 2 to a friend **to find the pattern in each of the 4 closed boxes.**

If he/she rolls the ball along **just one side:**

+ He/she might know that it **is** one box/es, by recognizing the pattern/s of some box/es,
+ But he/she also might know which box/es **are not** , by not being able to recognize the pattern/s of some box/es.

These are the boxes he/she is trying to find out about.



- Suppose he/she rolls the ball along the **right side** of all **closed** boxes.

If he does so, tell me, which box/es **does** he/she know about?

TICK (A ✓) (B) (C) (D) (none)

And, which boxes will seem the same, so he/she **can't** tell the difference?

TICK (A) (B) (C ✓) (D) (none)

- Suppose he/she now rolls the ball along the **bottom** of all **closed** boxes.

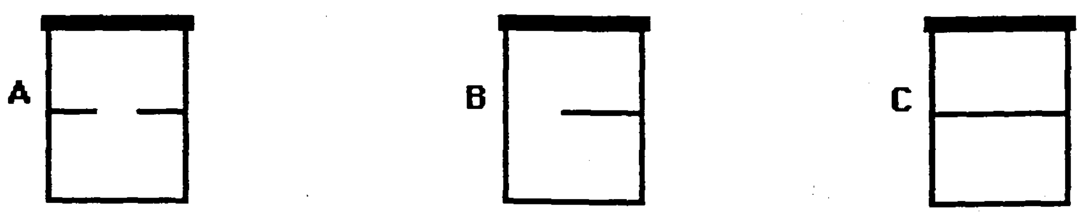
If he/she does so, which box/es **does** he/she know about?

TICK (A) (B) (C) (D) (none ✓)

And, which boxes will seem the same, so he/she **can't** tell the difference?

TICK (A) (B) (C) (D) (none ✓)

- Now, suppose you have these boxes:



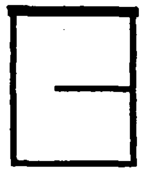
If you want to know which of the **closed** boxes is **B,**
What should be done or tested?

move the ball around the box

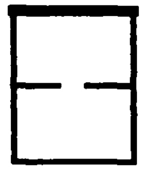
SALLY

Section 4:

Pick up the **white** box.
It has this pattern:
Try to move the ball around all the edges.



Now **try the same** with the **black** one.
It has this pattern:



Tell me, what is one thing the path of the ball depends on?

geting though the hole

What is a second thing the path of the ball depends on?

the cordors

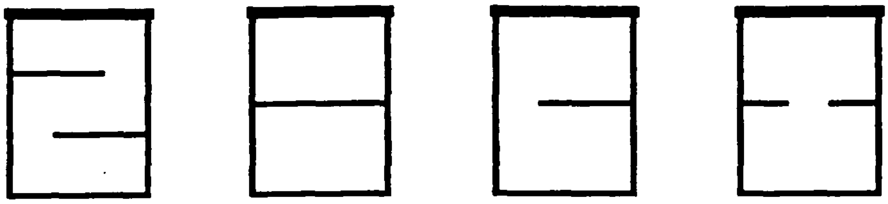
What is a third thing the path of the ball depends on?

how the Patten is

SALLY

Section 5:

You have these patterns in the boxes:



If you move the ball around all the edges in each box,

What is the **same or almost the same** about all these paths the ball can follow?

I don't no

What is another thing which is the same or almost the same about all these paths the ball can follow?

I don't no

Section 6:

You have the same boxes as in section 5.

If you move the ball around all the edges in each box,

Why can the ball follow different paths?

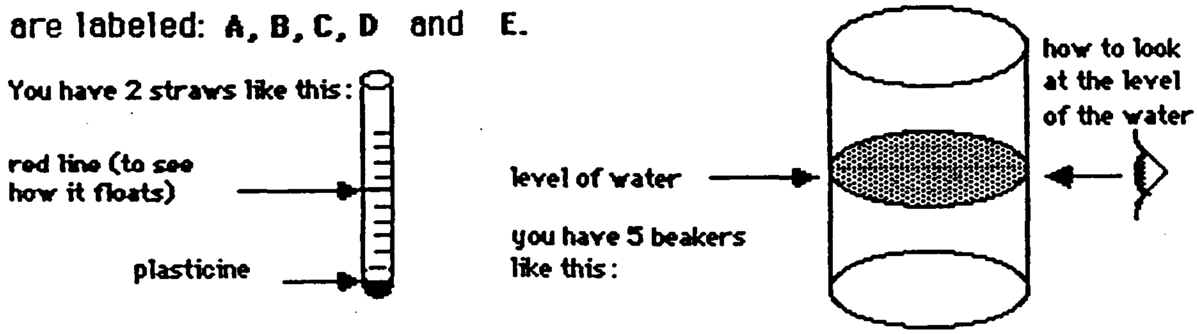
because the ball is near the ~~eage~~ edges

MARK

49 min

Section 1:

You have 2 straws: one is **yellow** and the other one is **green**.
You also have 5 beakers that contain the same amount of water. They are labeled: **A, B, C, D** and **E**.



Take the **yellow** straw. Put in 2 shots (from the ones that are in the bowl) and try to float it in all the beakers.

Look and see how it floats in each one.

Write what you noticed, here:

I noticed that in each one the straw would float about the same. I beacers A, B they would float the same but in beacers C, D and E they would float at just a ~~tino~~ tiney difference.

Take the **green** straw. Try floating it just in beaker **C**. Try it with more shots. Try it with fewer as well.

Look and see how it floats.

Write what you noticed, here:

I noticed when I put just 1 or 2 shots in it would float at mark 4 but when I put 6 and 10 shots in it would float at mark 5 or just above.

Take the yellow and the green straws. Fill the **yellow** with 4 shots and the **green** with 14 shots. Try to float them **together** in all the beakers.

Write what you noticed, here:

I noticed that the green straw would float at mark 6 but the yellow straw would float at mark 4 or 5.

MARK

Section 2:

You have been noticing how the straws float in all the beakers.
What do you think makes the straws float differently?

1 Weight makes a difference when floating the straws.	2 The amount of water makes a difference when floating the straws.
3 The kind of water makes a difference when floating the straws.	4 Put here anything else you think makes a difference.

Tick all the boxes you agree with:

1	✓	2		3	✓

Section 3:

Take the **green** straw and make its red line float **just at the level** of the water in beaker **C**, by putting in some shots.

What would be a good beaker to put the straw if you want its **red line to be up at the highest level** ?

Tick one box:

A		B		D		E	✓

Why do you think **that** is a good beaker to get the highest level of the red line?

I think it is a good beaker to put the straw in because when I tried to put the straw in before it did not float like it did in the others.

MARK

You have, again, the **green** straw with its red line just at the level of the water in beaker **C**.

What would be a good beaker to put the straw, if you want its **red line to be down at the lowest level** ?

Tick one box: A [ ] B [ ] D [✓] E [ ]

Why do you think **that** is a good beaker to get the lowest level of the red line?

I think it is a good beaker to put it in it went lower than it did in any other beacker

Section 4:

You have seen how straws behave when you were floating them.

What is one thing the floating level of the straws depends on?

It depends on how many shots you put in.

→What difference does it make?

The difference is that if you put alot of shots in it would go lower than it would if you put less shots in.

MARK

What is **another** thing the floating level of the straws depends on?

I don't think there is ~~saw~~ another thing that the level of the straw depends on.

→ What difference does it make?

I don't think there is another thing that the straw depends on.

Section 5:

You have noticed you need to do different things to make a straw float at a special mark.
Tell me, is there something which is the **same** about all the ways to make the straw float at a special mark?

No I don't think there is anything the same to make the straw float at a special mark.

Is there anything else which is the same about all the ways to make the straw float at a special mark?

No I don't think there is anything the same to make the straw float at a special mark.

MARK.

Section 6:

How do you think the water keeps the straw up?

I think the water Keeps the straw up because of the plastiene at the bottem

Why do you think only some things float, and others sink?

Because some things wight more than others. Because I think light things are easyer to float.

Could there be a liquid which makes everything float in it?

Tick one box: yes [ ] no [✓]

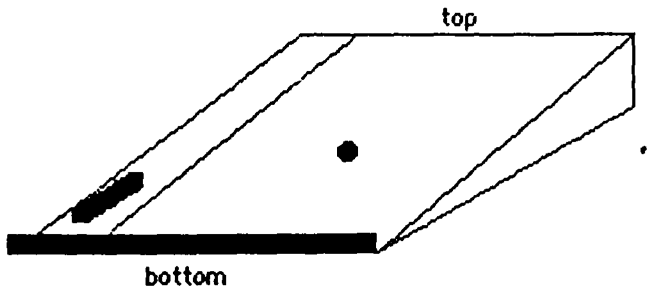
because... Because no matter what the liquid it depend on what it wigh's

July

42 min.

Section 1:

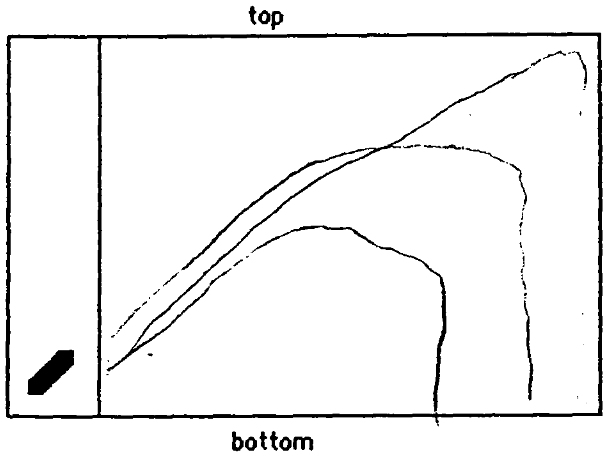
You have a sloping board and a ball you can make roll on it. Use the 'gun' to set the ball rolling. The stripe on the left is the firing area and you are not allowed to place the 'gun' out of there.



Roll the ball on the slope several times, firing up from the bottom, as shown.

Look to see how it goes.

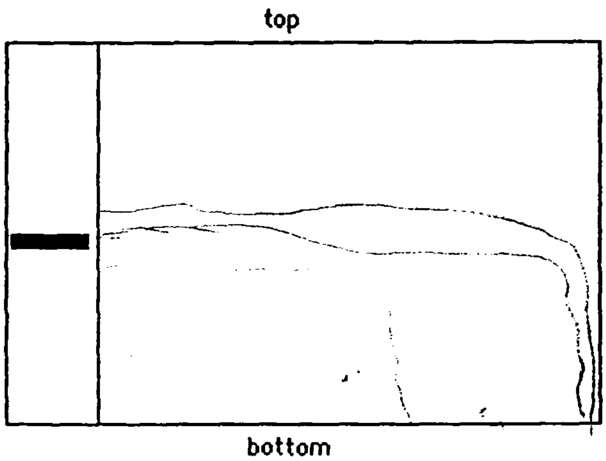
Draw here some of the different paths you see it takes.



Roll the ball on the slope several times, firing sideways from the middle, as shown.

Look to see hoe it goes.

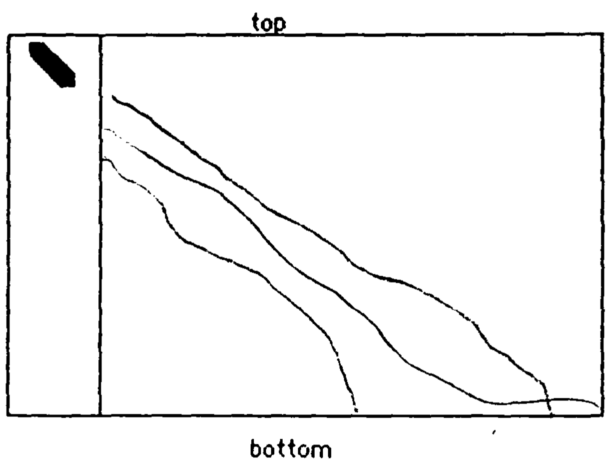
Draw here some of the different paths you see it takes.



Roll the ball on the slope several times, firing down from the top, as shown.

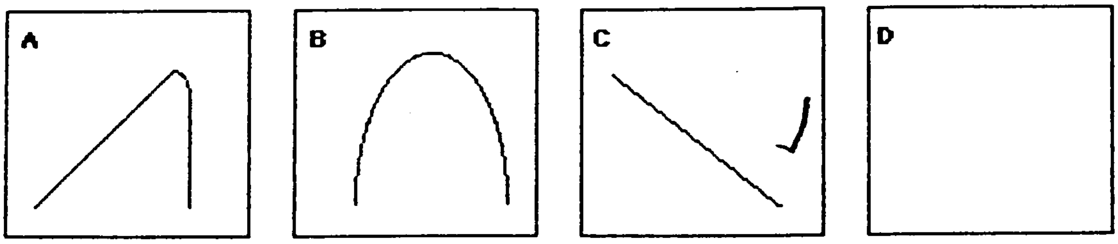
Look to see how it goes.

Draw here some of the different paths you see it takes.



Section 2:

You have been noticing the ball on the board.
Tell me, which of the following paths **are possible** ?



If you disagree with all of the other paths, draw the one you think is possible, here.

Tick the box/es you agree with:

A ✓ B C ✓

Why do you think the path/s you agree with is/are possible?

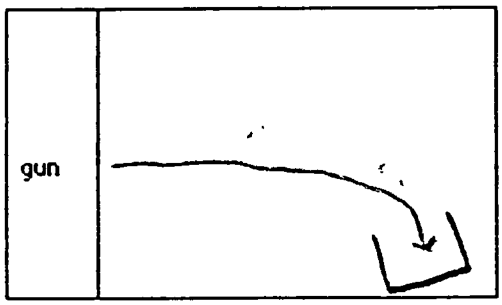
C is possible because it is a straight line. A is possible because it is like a straight line but when you get to the top you just have to turn the corner.

Section 3:

Suppose you have the gun placed as in the picture.
You also have a trap like a **U**

Which is a good place to put the **trap** if you want to get the ball in?

Draw here the trap and the possible path the ball will follow. →



Why do you think **that** is a good place to put the trap?

Because the gun is not that frow is away from the trap. It it like a straight line also that I think I could put the ball in there.

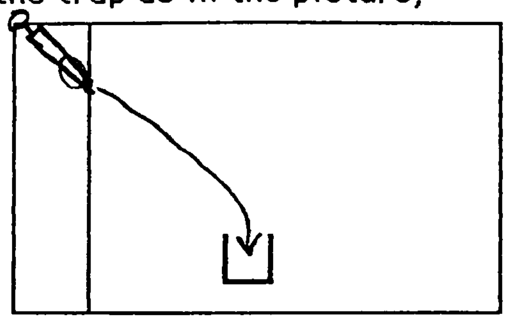
JULY

Suppose that we now fix the place of the trap as in the picture,

Which is a good place to put the **gun**

if you want to get the ball in?

Draw here the gun and the possible path the ball will follow. →

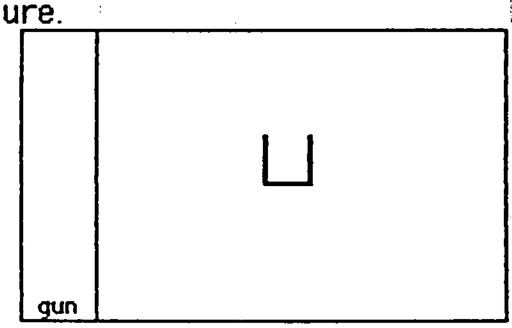


Why do you think **that** is a good place to put the gun?

Because ~~a~~ the trap is down the bottom and I think if I put the trap up the top that the ball will go in the trap

Section 4:

Set the gun and the trap as in the picture.



Try 5 times to get the ball into the trap.

As you probably noticed the ball can follow different paths.

What is one thing the path of the ball depends on?

~~The~~ If you don't move the gun you will be able to put the ball in the trap.

→ What difference does it make?

If you put the gun somewhere different the ball will not go in the trap. It will go down the bottom.

JULY

What is another thing the path of the ball depends on?

The path of the ball depends on where you put the trap and gun.

What difference does it make?

If you move the trap you will problaby find that you ~~to~~ will have to move the gun to. Then you can get the ball in.

Section 5:

By now you have probably seen a number of different paths.

Is there something which is the **same** about all of them?

Some of the lines are the same because some of ~~all~~ them are straight first. then they bend.

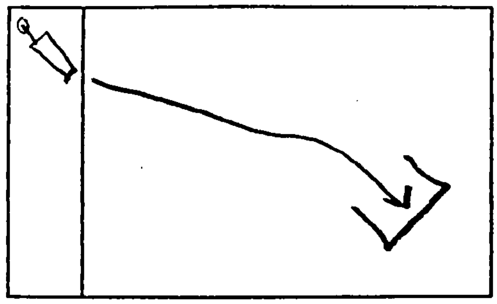
Is there anything else that is the **same** about all of them?

No.

JULY.

Section 6

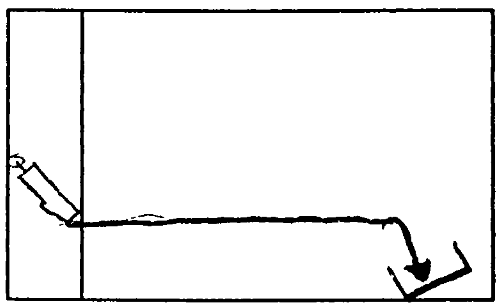
Draw here one of the paths you have seen before, when you were rolling the ball on the board.



Tell me, why does it go like that?

Because if you put the gun in the right path of the trap it would go into the trap.

Draw here another path.



Why do you think it goes like that?

If you could get the gun in a place that you think the ball will go in ~~to~~ do so.

APPENDIX C

INSTRUMENTS (Teacher's questionnaire).

Gary (5 years) YES. 9/10 years

1) YOU HAVE SEEN FOUR TASKS FOR CHILDREN AT THE END OF PRIMARY SCHOOL: ROLLER BALL, FLOATING, BALANCING AND BLACK BOXES.

• WHAT PERCENTAGE OF CHILDREN DO YOU THINK WOULD **DO WELL, NOT DO WELL, OR BE IN BETWEEN.** (Percentages should add 100 in each row)

	%	DO WELL	%	IN BETWEEN	%	NOT DO WELL
- ROLLER BALL	30	Get the ball inside the trap.	50	The ball hitting the trap or near by; without going into it.	20	The ball passing far away from the trap.
- FLOATING	20	Work out the exact amount of shot to the nearest one shot.	50	Get the amount correct to between 2-4 shots.	30	Only get the amount that is wrong by 5 or or more shots.
- BALANCING	50	Give a correct selection of bricks.	40	Give an almost correct selection of bricks.	10	Give a totally wrong selection of bricks.
- BLACK BOXES	80	Select the exact pattern that matches the one in the box.	15	Select one pattern that is quite similar to the one in the box.	5	Select one pattern that is not very similar to the one in the box.

• WOULD YOU USE THESE TASKS AS TEACHING DEVICES?

	YES, ALL	YES, SEVERAL	MAYBE ONE	NO, NONE
Tick one box	(crossed out)	✓		

because... practical activities such as these would be important in science work.

2) THESE TASKS INVOLVE CHILDREN IN SOME REASONING.
HOW DIFFICULT DO YOU THINK THIS KIND OF REASONING IS FOR CHILDREN OF 10/11 YEARS?

	VERY DIFFICULT	FAIRLY DIFFICULT	FAIRLY EASY	EASY
Tick one box		✓		

Gary

3) BELOW ARE SOME PROBLEM SOLVING PROCESSES. IN THE TASKS YOU HAVE SEEN WHAT PERCENTAGE OF CHILDREN OF 10-11 DO YOU THINK WOULD PERFORM EACH PROCESS **WELL, POORLY** OR **IN BETWEEN.** (Make sure each row adds 100)

	% WELL	% IN BETWEEN	% POORLY.
- NOTICING CHANGES, DIFFERENCES.	85	10	5
- SETTING THE RIGHT CONDITIONS FOR SOMETHING TO HAPPEN.	30	60	10
- IDENTIFYING FACTORS AFFECTING SOME EVENTS.	40	50	10
- MAKING GENERALIZATIONS ABOUT PATTERNS OF BEHAVIOUR.	20	50	30
- IMAGINING AGENTS RESPONSIBLE FOR SOME PHENOMENON.	10	10	80

4) TO WHAT EXTENT DO YOU THINK THAT THE PROCESSES BELOW ARE ALSO IMPORTANT EDUCATIONAL AIMS? (Tick one box in each row)

	VERY	FAIRLY	NOT VERY	DEFINITELY NOT
- BEING SYSTEMATIC.	✓			
- NOTICING CHANGES, DIFFERENCES		✓		
- FORECASTING EVENTS.		✓		
- SETTING THE RIGHT CONDITIONS FOR THINGS TO HAPPEN.		✓		
- MAKING GENERALIZATIONS ABOUT PATTERNS OF BEHAVIOUR.	✓			
- PLANNING SEARCHES.	✓			
- IMAGINING AGENTS RESPONSIBLE FOR PHENOMENA.		✓		
- MAKING RELEVANT NOTES.		✓		
- IDENTIFYING FACTORS AFFECTING EVENTS.	✓			

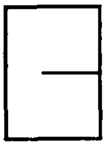
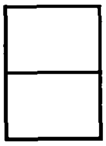
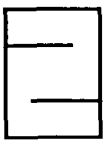
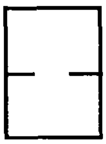
APPENDIX D

SEARCH-SPACE TABLES.

Distance	Groups	Config.	Succ.	Fail.	%	% within a group	Children*	Children* (group)	Avge.	Range
			Attempts							
Same distance	Single	Identical	37		9.46	25.00	10	20	3.70	1-13
		Equivalent	93		23.78	62.83	20		4.65	1-14
		Different		18	4.60	12.16	6		3.00	1-06
	Double	Identical	5		1.27	5.10	4	15	1.25	1-02
		Equivalent	39		9.97	39.79	11		3.54	1-09
		Compensated	16		4.09	16.32	6		2.66	1-05
		Different		38	9.71	38.77	10		3.80	1-07
Different distance	Pairwise	Compensated	26		6.64	27.65	10	16	2.60	1-09
		Crossed		18	4.60	19.14	8		2.25	1-05
		Non-compensated		20	5.11	21.27	8		2.50	1-06
		Not-enogh comp.		30	7.67	31.91	8		3.75	1-12
	Combina-torial	Appropriate	17		4.34	33.33	10	14	1.70	1-04
		Not-appropriate		34	8.69	66.66	11		3.09	1-08
			233	158	99.93					

* The second column shows the number of children that performed at least in one of the configurations belonging to the groups: single, double, pairwise and combinatorial.

Table 1: number of attempts made and children involved in each configuration and group of configurations in the BALANCING task.

Attempts									
Groups	Configurations		Num.	%	% within a box	Children*		Avge.	Range
		Left	83	7.22	35.31	22		3.77	1-16
C		Right	89	7.75	37.87	23	24	3.86	1-13
		Obstacles	63	5.48	26.80	22		2.86	1-09
		Left	103	8.97	27.32	19		5.42	1-15
B		Right	98	8.53	25.99	21	24	4.66	1-12
		Obstacles	176	15.33	46.68	24		7.33	1-17
		Left	76	6.62	36.53	19		4.00	1-18
Z		Right	77	6.70	37.01	24	24	3.20	1-10
		Obstacles	55	4.79	26.44	20		2.75	1-09
		Left	103	8.97	31.40	22		4.68	1-12
I		Right	97	8.44	29.57	19	24	5.10	1-11
		Obstacles	128	11.14	39.02	23		5.56	1-13
			1148	99.94					

* Children's second column shows the number of them that made attempts in each box.

Table 2: number of attempts made and children involved in each configuration and group of configurations when looking within each box in the BLACK BOXES task.

Groups	Config.	Test	Not test	%*	% within a group	Children**		Avge.	Range
		Attempts							
C/B (r,l)	Same	1 (l)	-	0.08	7.14	1	12	1.00	---
	Different		13	1.13	92.85	12		1.08	1-2
C/Z (r)	Same	4 (2l,2r)	1	0.43	25.00	4	14	1.25	1-2
	Different		15	1.30	75.00	11		1.36	1-2
C/I (r,l)	Same	7 (4r,3ob)	-	0.60	29.16	6	15	1.16	1-2
	Different		17	1.48	70.83	11		1.54	1-3
B/Z (r)	Same	5 (4r,1ob)	1	0.52	28.57	5	16	1.20	1-2
	Different		15	1.30	71.42	14		1.07	1-2
B/I (r,l,o) ***	Same	6 (3l,3r)	-	0.52	30.00	4	12	1.50	1-2
	Different		14	1.21	70.00	10		1.40	1-2
Z/I (r)	Same	5 (4l,1r)	4	0.78	42.85	5	13	1.80	1-2
	Different		12	1.04	57.14	9		1.33	1-2
		28	92						

* Based on the total amount of attempts (1148).
** The second column indicates the number of children that changed box as indicated by the above pair of coloured boxes.
*** r, l, o, stands for 'right', 'left' and 'obstacles' sides of the boxes.

Table 3: number of attempts made and children involved in each configuration and group of of configurations when looking at comparing different boxes in the BLACK BOXES task.

Groups	Config.	Attempts Num.	%	% within a group	Children*		Avge.	Range
Solution A	Far	60	11.78	65.21	20		3.00	1-11
	Very near	11	2.16	11.95	09	24	1.22	1-02
	Exact	21	4.12	22.82	16		1.21	1-02
Solution B	Far	38	7.46	48.71	19		2.00	1-04
	Very near	15	2.94	19.23	09	24	1.66	1-05
	Exact	25	4.91	32.05	16		1.56	1-03
Solution C	Far	48	9.43	61.53	20		2.40	1-06
	Very near	14	2.75	17.94	08	24	1.75	1-04
	Exact	16	3.14	20.51	12		1.33	1-02
Solution D	Far	80	15.71	73.39	23		3.47	1-12
	Very near	17	3.33	15.59	13	24	1.30	1-02
	Exact	12	2.35	11.00	09		1.33	1-02
Solution E	Far	118	23.18	77.63	23		5.13	1-12
	Very near	18	3.53	11.84	14	24	1.28	1-02
	Exact	16	3.14	10.52	12		1.33	1-02
		509	99.93					

* Children's second column shows the number of them that make attempts in each solution or beaker.

Table 4: number of attempts made and children involved in each configuration and group of configurations when performing the FLOATING task.

Groups	Config.*	Attempts Num.**		Succ.	%	% in a group	Children***		Avge.	Range
Horizontal at top	GuTulf	10	219	10	0.62	13.72	4	16	2.50	1-05
	GuTulu	20		8	1.25		2		10.00	7-13
	GuTurf	31		15	1.94		6		5.16	1-09
	GuTuru	158		28	9.89		12		13.16	3-38
Downwards	GuTdlf	21	569	13	1.31	35.65	4	21	5.25	1-10
	GuTdlu	37		15	2.31		8		4.62	1-12
	GuTdrf	171		75	10.71		8		21.37	2-64
	GuTdru	340		106	21.30		17		20.00	1-92
Upwards	GdTulf	66	295	22	4.13	18.48	8	18	8.25	1-21
	GdTulu	23		5	1.44		5		4.60	1-13
	GdTurf	97		29	6.07		10		9.70	3-23
	GdTuru	109		18	6.82		12		9.08	3-28
Horizontal at bottom	GdTdlf	43	513	20	2.69	32.14	5	22	8.60	1-21
	GdTdlu	62		15	3.88		7		8.85	4-15
	GdTdrf	123		39	7.70		10		12.30	2-40
	GdTdru	285		67	17.85		18		15.83	1-68
			1596	485	99.91	99.99				

* The coding is as follows: G = gun; T = trap; u = up; d = down; l = left; r = right; u (in the far right) = upwards; f = facing.
** Second set of data shows number of attempts made in each group of configurations.
*** Children's second column shows the number of children that performed in each group of configurations.

Table 5: number of attempts and children involved in each configuration and group of configurations when performing the ROLLING task.

APPENDIX E

CATEGORIES AND RESULTS.

Table 1: categories used to group children's answers and number of children in each category for the BALANCING task per 'process'.

NOTICING		
a) Opposite stripe=level; add on non-opposite= tilted		
-Noticing balance tilted, plus some kind of reason (not given by senses):		13
- Noticing balance tilted:		11
- Not noticing the balance tilted:		0
	TOTAL	24
b) Non-opposite stripe=tilted; add on opposite= level.		
- Noticing balance level, plus some kind of reason:		7
- Noticing balance level:		16
-Noticing a change:		1
- Not noticing balance level:		0
	TOTAL	24

UNDERSTANDING		
a) What makes the balance level?		
- Same amount of plastic:		2
- Same size:		4
- Same amount of dots:		3
- Same size & amount of dots:		8
- Same size & amount of plastic:		6
- Same amount of dots & plastic:		1
- Same size, amount of dots and plastic:		0
	TOTAL	24
- Continuous:		2
- Continuous & discrete:		7
- Discrete:		15
	TOTAL	24
- Continuous:		2
- Discrete or affected:		4
- Qualitative or affected:		10
- Discrete and qualitative:		8

WHAT-IF REASONING		
a) Choose weight being distance constant.		
- Half the weight than the other side (c):		10
- Less weight than in the other side:		1
- Same weight as in the other side:		11
- More weight than in the other side:		2
	TOTAL	24
Reasons:		
- A plausible justification:		7
- Justification based on past experience:		4
- Not a plausible justification:		13
- Not a clear justification:		0
	TOTAL	24
b) Choose distances being weight constant.		
- Taking into account distance:		16
- Not taking into account distance (just weight in the other side):		8
	TOTAL	24
Reasons:		
- A plausible justification:		12
-Justification based on past experience:		0
- Not a plausible justification:		12
- Not a clear justification:		0
- Not a justification given:		0
	TOTAL	24

IDENTIFYING VARIABLES		
a) One thing.		
- Plausible variable:		23
- Not a plausible variable:		1
	TOTAL	24
Reason:		
- A plausible difference:		21
- Not a plausible difference:		2
- Not a clear difference:		1
	TOTAL	24
b) Second thing.		
- Plausible variable:		11
- Not a plausible variable:		3
- Not identified: (A: 2 already identified)		10
	TOTAL	24
Reasons:		
- A plausible difference:		9
- Not a plausible difference:		5
- Not a clear difference:		0
- Not a reason given: (A: 2 based repeated var)		10
	TOTAL	24

MAKING GENERALIZATIONS		
a) One thing that is the same:		
- A sound generalization:		12
- Not a sound generalization:		7
- Not generalization made:		5
	TOTAL	24
b) Another thing that is the same:		
- A sound generalization:		3
- Not a sound generalization:		3
- Not generalization made:		18
	TOTAL	24

IMAGINING CAUSES		
a) Same effect, tilting, by using different means.		
- Explained in terms of something acting down:		5
- Explained in terms of having the same property (heaviness):		8
- Explained in terms of the apparatus (visible):		11
	TOTAL	24
b) Same effect, balance level, by using different means.		
- Explained in terms of something acting down:		4
- Explained in terms of having the same property (heaviness):		13
- Explained in terms of the apparatus: (visible):		5
- Not a clear reason:		1
- Not explanation given:		1
	TOTAL	24

Table 2: categories used to group children's answers and number of children in each category for the BLACK BOXES task per 'process'.

NOTICING		
a) Noticing a close difference in distance.		
- Noticing the correct difference:		14
- Noticing an incorrect difference:		1
- Not noticing the difference:		9
	TOTAL	24
b) Noticing a fair difference in distance.		
- Noticing the correct difference:		20
- Noticing an incorrect difference:		1
- Not noticing the difference:		3
	TOTAL	24

UNDERSTANDING (STRATEGY)		
a) Best way to do it.		
- Stick with a box:		16
- Choose a test:		2
- Stick with a pattern:		6
	TOTAL	24

WHAT-IF REASONING		
a) Set the ball in all four corners.		
- 4/4:		12
- 3/4:		9
- 2/4:		3
- 1/4:		0
- 0/4:		0
	TOTAL	24
b) Reasoning.		
b1) What happens if they roll the ball along the right side.		
* Positive discrimination:		
- Identification of positive discrimination:		15
- Partial identification of positive discrim.:		2
- Not identification of positive discrim.:		2
- Faulty logic		5
	TOTAL	24
* Negative discrimination:		
- Identification of lack of discrimination:		14
- Partial identification of lack of discrim.:		4
-Not identification of lack of discrim.:		1
- Faulty logic:		5
	TOTAL	24

b2) What happens if they roll the ball along the bottom.		
* Identif. of lack of discrim. in negative mode:		
- Identification of lack of discrimination:		18
- Not identification of lack of discrimination:		5
- Faulty logic		1
	TOTAL	24
* Identification of lack of discrimination in a positive mode.		
- Identification of lack of discrimination:		17
- Partial identification of lack of discrimination:		5
- Not identification of lack of discrimination:		1
- Faulty logic:		1
	TOTAL	24
c) What should be done or tested:		
- Test needed:		10
- Different tests:		6
- Move the ball around all the edges:		6
- Not clear:		1
- Not answered:		1
	TOTAL	24

IDENTIFYING VARIABLES		
a) One thing.		
- Plausible variable:		12
- Not a plausible variable:		12
- Not a clear variable:		0
- Not identified:		0
	TOTAL	24
b) Second thing.		
- Plausible variable:		6
- Not a plausible variable:		10
- Not a clear variable:		0
- Not identified:		8
	TOTAL	24
c) Third thing.		
- Plausible variable:		3
- Not a plausible variable:		6
- Not a clear variable:		1
- Not identified:		14
	TOTAL	24

MAKING GENERALIZATIONS		
a) One thing that is the same.		
- Sound generalization:		15
- Not a sound generalization:		5
- Not a generalization made:		2
- Negation of any generalization:		2
	TOTAL	24
b) Second thing that is the same.		
- Sound generalization:		9
- Not a sound generalization:		7
- Not a generalization made:		4
- Negation of any generalization:		4
	TOTAL	24

IMAGINING CAUSES		
a) Why can the ball follow different paths?.		
- Plausible agent of causation:		17
- Not a plausible agent of causation:		
External agent:		5
Causation based on actions:		2
	TOTAL	24

Table 3: categories used to group children's answers and number of children in each category for the FLOATING task per 'process'.

NOTICING.		
a) Weight is constant & concentration changes.		
- Notice different floating levels with some detail:		12
- Notice different floating levels without detail:		8
- Not mention of different floating levels:		4
	TOTAL	24
b) Concentration is constant & weight changes.		
- Noticing different floating levels in relation to the weight used:		22
- Noticing different floating levels:		2
- Not mention different floating levels:		0
	TOTAL	24
c) Weight is constant & concentration changes. (2 straws, together; different weights)		
- Noticing both, the effect of different concentrations and diff. masses:		0
- Noticing the effect of diff. concentrations:		9
- Noticing the effect of diff. masses:		9
- Not noticing clearly either one effect neither the other:		6
	TOTAL	24

UNDERSTANDING.		
a) What makes the straw float differently?		
- Weight:		4
- Amount of water:		1
- Kind of water:		1
- Weight & amount of water:		4
- Weight & kind of water:		12
- Amount of water & kind of water:		0
- Weight, amount of water & kind of water:		2
	TOTAL	24
- Water playing a role:		
Δ Kind of water:		13
Δ Amount of water:		5
Δ Amount & kind of water:		2
- Water not playing a role:		4
	TOTAL	24
- Weight playing a role:		
Δ Weight alone:		4
Δ W & amount of water:		4
Δ W & kind of water:		12
Δ W, amount & kind of water		2
- Weight not playing a role:		2
	TOTAL	24
- Amount of water interfering		
Δ Weight & kind of water:		12
Δ Weight or kind of water:		5
Δ Amount of water:		7
	TOTAL	24

WHAT-IF REASONING		
a) Highest level.		
- Correct direction & exact beaker:		15
- Correct direction & inexact beaker:		5
- Incorrect direction:		4
	TOTAL	24
Reason:		
- A plausible justification:		5
- Not a plausible justification:		6
- Justification based on past experience:		12
- Not a reason given:		1
	TOTAL	24
b) Lowest level:		
- Correct direction & exact beaker:		16
- Correct direction & inexact beaker:		0
- Incorrect direction:		8
	TOTAL	24
Reason:		
- A plausible justification:		5
- Not a plausible justification:		7
- Justification based on past experience:		11
- Not a clear reason:		1
	TOTAL	24

IDENTIFYING VARIABLES		
a) One thing.		
- Plausible variable:		22
- Not a plausible variable:		2
	TOTAL	24
Reason:		
- A plausible difference:		20
- Not a plausible difference:		4
	TOTAL	24
b) Second thing.		
- Plausible variable:		16
- Not a plausible variable:		4
- No variable identified:		4
	TOTAL	24
Reason:		
- A plausible difference:		13
- Not a plausible difference:		7
- Not difference identified:		4
	TOTAL	24

MAKING GENERALIZATIONS.		
a) One thing that is the same.		
- Sound generalization:		8
- Not a sound generalization:		7
- No generalization made:		9
	TOTAL	24
b) Another thing that is the same.		
- Sound generalization:		2
- Not a sound generalization:		2
- No generalization made:		20
	TOTAL	24

IMAGINING CAUSES		
a) How water keeps the straw up.		
- No role assigned to water:		13
- Role assigned to water:		11
	TOTAL	24
b) Why things float or sink.		
- In terms of heaviness or lightness:		21
- Factor other than heaviness or lightness:		1
- In terms of more than one factor:		2
	TOTAL	24
c) Liquid that makes everything float in it.		
• Yes:	SUB-TOTAL	12
Reasons:		
- In terms of 'hardness':		8
- In terms of something added:		0
- In terms of making things light:		3
- Not a clear reason:		1
	SUB-TOTAL	12

Table 4: categories used to group children's answers and number of children in each category for the ROLLING task per 'process'.

NOTICING		
a) Firing from bottom to top.		
- Noticing plausible paths:		11
- Noticing a combination of plausible & not plausible paths:		8
- Noticing not-plausible paths:		5
	TOTAL	24
b) Firing sideways from the middle.		
- Noticing plausible paths: (OL: 3 included straight ones) (A: 1 included straight ones)		6
- Noticing a combination of plausible & not-plausible paths (OL: 2 included straight paths) (A: 1 included straight paths)		15
- Noticing not-plausible paths:		3
	TOTAL	24
c) Firing from top to bottom:		
- Noticing plausible paths: (OL: 7 included straight ones) (A: 5 included straight paths)		9
- Noticing plausible & not-plausible paths: (OL: 2 included straight paths) (A: 3 included straight paths)		13
- Noticing not-plausible paths:		2
	TOTAL	24

UNDERSTANDING.		
a) Which paths are possible?		
- Parabola:		1
- Dropping:		2
- Straight:		7
- Dropping & straight:		7
- Parabola & straight:		5
- Dropping, parabola & straight:		2
	TOTAL	24
Rearrangement of categories:		
- Possible:		1
- Possible & not-possible:		7
- Not-possible:		16
	TOTAL	24
Another arrangement of categories:		
-Plausible:		13
- Plausible and non-plausible:		9
- Non plausible:		2
	TOTAL	24
Reasons:		
- External reasons:		5
- Graphical or motor description:		13
- Past experience:		4
- Opposition to other path:		1
- What is expected:		1
	TOTAL	24

WHAT-IF REASONING		
a) Good place to put the trap (gun at the middle).		
- Bottom:		22
- Middle:		2
- Top:		0
	TOTAL	24
Prediction:		
- Parabola:		6
- Horizontal (curved):		4
- Straight line (diagonal):		5
- Not possible:		9
	TOTAL	24
Reasons:		
- A plausible justification:		7
- Past experience:		8
- Not a plausible justification:		9
	TOTAL	24
b) Good place to put the gun (trap at middle bottom)		
- At the level of the trap:		6
- Above-middle:		7
- Above-top:		11
	TOTAL	24
Predictions:		
- Parabola:		6
- Horizontal (curved):		4
- Straight line (diagonal):		2
- Not possible:		12
	TOTAL	24

Reasons:		
- External reasons:		8
- Past experience:		2
- Graphical or motor description:		14
- Functional:	TOTAL	24

IDENTIFYING VARIABLES a) One thing.		
- Plausible variable:		20
- Not a plausible variable:		4
- Not identified:		0
	TOTAL	24
Reasons:		
- Plausible difference:		9
- Not a plausible difference:		15
- Not a clear reason:		0
- Not a reason given:		0
	TOTAL	24
b) Second thing.		
- Plausible variable: how 'releasing' is made; OL: 1 & A: 1.		12
- Not a plausible variable:		8
- Not a clear variable:		0
- Not identified: (OL: one was the same as before)		4
	TOTAL	24
Reasons:		
- Plausible difference:		7
- Not a plausible difference:		13
- Not a clear reason:		0
- Not a reason given: (OL: one was given before)		4
	TOTAL	24

MAKING GENERALIZATIONS		
a) One thing that is the same.		
- A sound generalization:		13
- Not a sound generalization:		5
- Not a generalization made:		0
- Negation of a generalization:		6
	TOTAL	24
b) Another thing that is the same:		
- A sound generalization:		7
- Not a sound generalization:		3
- Not a generalization made:		8
- Negation of a generalization:		6
	TOTAL	24

IMAGINING CAUSES		
a) Cause for 1st drawing.		
- Based on external conditions:		4
- Based on objectified actions:		11
- Based on actions made:		7
- Artificialism:		1
- Not identified:		1
	TOTAL	24
b) Cause for 2nd drawing.		
- Based on external conditions:		4
- Based on objectified actions:		10
- Based on actions made:		10
- Not identified:		0
	TOTAL	24

APPENDIX F

CORRELATION TABLES FOR ALL QUESTIONS. (Structured tasks).

	N1	N2	U	WR1	WR2	WR12	WR3	WR4	WR34	V1	V2	V12	V3	V4	V34	G1	G2	C1	C2
N1	1																		
N2	0.603	1																	
U	0.005	0.155	1																
WR1	-0.21	-0.43	0.025	1															
WR2	-0.17	-0.41	0.129	0.794	1														
WR12	-0.26	-0.38	0.203	0.858	0.949	1													
WR3	0.151	0.041	0.15	0.521	0.662	0.608	1												
WR4	-0.08	0	-0.07	0.558	0.569	0.588	0.775	1											
WR34	-0.08	0	-0.07	0.558	0.569	0.588	0.775	1	1										
V1	-0.19	-0.3	0.148	-0.24	-0.3	-0.33	-0.16	-0.21	-0.21	1									
V2	0.158	-0.06	0.073	0.081	0.036	0	0.228	0.126	0.126	0.552	1								
V12	0.158	-0.06	0.073	0.081	0.036	0	0.228	0.126	0.126	0.552	1	1							
V3	0.251	0	0.194	0.043	0.38	0.294	0.258	0.167	0.167	-0.21	-0.13	-0.13	1						
V4	0.194	-0.04	0.251	0.277	0.613	0.506	0.6	0.43	0.43	-0.27	0.033	0.033	0.775	1					
V34	0.194	-0.04	0.251	0.277	0.613	0.506	0.6	0.43	0.43	-0.27	0.033	0.033	0.775	1	1				
G1	0.251	0	0.065	0.386	0.38	0.392	0.258	0.333	0.333	-0.21	0.126	0.126	0.167	0.258	0.258	1			
G2	-0.16	-0.18	0.122	0.438	0.538	0.593	0.293	0.378	0.378	-0.55	-0.24	-0.24	0.378	0.488	0.488	0.378	1		
C1	-0.19	-0.36	-0.1	0.069	0.275	0.189	0.194	0.107	0.107	-0.07	-0.12	-0.12	0.214	0.36	0.36	-0.21	0.446	1	
C2	-0.1	-0.12	-0.11	0.468	0.483	0.476	0.564	0.606	0.606	-0.05	0.275	0.275	0.243	0.438	0.438	0.364	0.458	0.078	1

Table 1: Pearson Product-Moment Correlations for all questions in the BALANCING task.
N = Noticing; U = Understanding; WR = What-if reasoning; V= Identifying variables; G= Making generalizations; C = Imagining causes
(The numbers indicate the order of the question in the instrument; the order of the processes here corresponds to their order in the instrument).

	N1	N2	U	P	HR1	HR2	HR3	HR4	Dt	V1	V2	V3	G1	G2	C
N1	1														
N2	0.076	1													
u	-0.32	0.145	1												
p	0.091	0.241	0.174	1											
HR1	-0.1	0.411	0.198	0.403	1										
HR2	-0.07	0.307	0.178	0.493	0.721	1									
HR3	-0.23	0.321	0.365	0.129	0.565	0.322	1								
HR4	-0.23	0.618	0.319	0.225	0.766	0.542	0.668	1							
Dt	-0.39	0.258	0.149	0.138	0.471	0.283	0.474	0.38	1						
V1	0	0.224	0.065	-0.18	0.204	0.098	0.308	0.329	0.481	1					
V2	-0.1	0.258	0.187	-0.04	-0.47	-0.34	-0.18	-0.11	-0.11	0	1				
V3	0.319	-0.17	-0.07	0.158	-0.08	0	0.194	-0.17	-0.15	-0.13	-0.22	1			
G1	0.044	0.346	0.017	0.046	0.158	0	0.026	0.357	-0.3	-0.26	0.05	-0.23	1		
G2	-0.04	0.115	0.117	0.077	0.474	0.203	0.185	0.323	0	-0.09	-0.45	-0.29	0.6	1	
C	-0.17	0.205	0.314	0.082	-0.06	0.036	0.122	0.199	0.318	0.275	0.37	-0.31	0.071	-0.07	1

Table 2: Pearson Product-Moment Correlations for all questions in the BLACK BOXES task.

N = Noticing; U = Understanding; P = Practically setting; HR = Hypothetical reasoning; Dt = Devising test; V = Identifying variables; G = Making generalizations; C = Imagining causes (The numbers indicate the order of the question in the instrument; the order of the processes here corresponds to their order in the instrument).

	N1	N2	N3	U	WR1	WR2	WR12	WR3	WR4	WR34	V1	V2	V12	V3	V4	V34	G1	G2	C1	C2	C3
N1	1																				
N2	0.14	1																			
N3	-0.26	0.17	1																		
U	0.2	0.05	0.2	1																	
WR1	0.12	0.17	0.32	0.25	1																
WR2	0.21	0.18	0.07	0.02	0.77	1															
WR12	0.21	0.18	0.07	0.02	0.77	1	1														
WR3	-0.04	0.11	0.41	0.21	0.62	0.55	0.55	1													
WR4	0.18	0.14	0	0.1	0.64	0.77	0.77	0.55	1												
WR34	0.07	0.1	0.19	0.11	0.6	0.73	0.73	0.68	0.94	1											
V1	-0.07	0.46	0.17	0.23	0.56	0.39	0.39	0.43	0.34	0.29	1										
V2	0.2	0.27	0	0.34	0.54	0.58	0.58	0.16	0.5	0.43	0.67	1									
V12	0.2	0.27	0	0.34	0.54	0.58	0.58	0.16	0.5	0.43	0.67	1	1								
V3	-0.16	0.11	0.61	0.53	0.27	0.17	0.17	0.25	0.32	0.46	0.11	0.4	0.4	1							
V4	0	0.3	0.39	0.35	0.44	0.36	0.36	0.53	0.56	0.64	0.3	0.45	0.45	0.71	1						
V34	0	0.3	0.39	0.35	0.44	0.36	0.36	0.53	0.56	0.64	0.3	0.45	0.45	0.71	1	1					
G1	0.4	0.21	0.41	0.32	0.43	0.46	0.46	0.31	0.16	0.23	0.21	0.32	0.32	0.31	0.35	0.35	1				
G2	0.07	0.09	-0.17	0.32	0.03	0.04	0.04	0.21	0.07	0.1	0.09	0.14	0.14	0.21	0.3	0.3	0.11	1			
C1	-0.19	-0.03	-0.05	0.16	-0.06	0.11	0.11	0.3	0.32	0.4	0.28	0.19	0.19	0.12	0.08	0.08	-0.3	0.03	1		
C2	0.27	0.09	0.17	0.32	0.03	0.25	0.25	0.21	0.27	0.29	0.09	0.14	0.14	0.21	0.3	0.3	0.43	-0.09	0.33	1	
C3	-0.3	-0.03	0.34	0.16	0.05	0.11	0.11	0.12	-0.02	0.08	-0.03	-0.04	-0.04	0.3	-0.08	-0.08	-0.12	0.03	0.5	0.03	1

Table 3: Pearson Product-Moment Correlations for all questions in the FLOATING task.
N = Noticing; U = Understanding; WR = What-if reasoning; V = Identifying variables; G = Making generalizations; C = Imagining causes
(The numbers indicate the order of the question in the instrument; the order of the processes here corresponds to their order in the instrument).

	N1	N2	N3	U1	U2	U3	WR2	WR3	WR23	WR5	WR6	WR56	V1	V2	V12	V3	V4	V34	G1	G2	C1	C2
N1	1																					
N2	0.56	1																				
N3	0.2	0.47	1																			
U1	0.26	0.23	-0.2	1																		
U2	0.19	0.28	0.4	0.33	1																	
U3	0.43	0.19	0.08	0.07	0.37	1																
WR2	0.41	0.3	0.36	0.14	0.04	0.02	1															
WR3	-0.1	0.11	0.05	-0.39	-0.41	-0.12	-0.11	1														
WR23	0.17	0.16	0.17	-0.22	-0.37	-0.1	0.57	0.58	1													
WR5	0.31	0.53	0.4	0.14	-0.01	-0.48	0.25	0.1	0.3	1												
WR6	0.32	0.36	0.13	0.1	-0.02	0.14	-0.04	0.42	0.24	0.17	1											
WR56	0.39	0.53	0.1	0.32	-0.05	-0.33	0.16	0.39	0.34	0.57	0.68	1										
V1	0.14	0.47	0.03	0.1	0.32	-0.09	0.02	0.09	0.13	0.43	0.12	0.25	1									
V2	-0.14	-0.02	-0.09	-0.21	0.12	0.08	-0.23	0.08	-0.02	-0.2	0.02	-0.2	0.35	1								
V12	-0.14	-0.02	-0.09	-0.21	0.12	0.08	-0.23	0.08	-0.02	-0.2	0.02	-0.2	0.35	1	1							
V3	0.11	0.21	0.21	0.07	0.07	-0.21	0.33	0.1	0.18	0.26	0.18	0.33	0.22	-0.26	-0.26	1						
V4	-0.09	0.02	0.14	-0.26	-0.03	-0.35	0.07	0.18	0.1	0.15	0.27	0.25	0.29	0.07	0.07	0.64	1					
V34	-0.09	0.02	0.14	-0.26	-0.03	-0.35	0.07	0.18	0.1	0.15	0.27	0.25	0.29	0.07	0.07	0.64	1	1				
G1	0.08	-0.09	0.03	0.02	0.14	-0.06	-0.05	-0.2	-0.23	0.01	-0.16	-0.16	0.04	-0.15	-0.15	0.08	0.04	0.04	1			
G2	0.03	0.17	0.29	-0.43	0.11	-0.12	-0.03	0.18	0.1	0.44	0.07	0.01	0.29	0.07	0.07	0.09	0.19	0.19	0.22	1		
C1	-0.13	0.16	-0.05	-0.01	0.12	0.14	-0.05	0.26	0.02	-0.09	0.18	0.01	0.19	0.23	0.23	-0.06	-0.07	-0.07	0.2	0.45	1	
C2	-0.04	-0.02	0.17	0.03	-0.2	-0.18	-0.05	0.25	0.14	0.27	-0.03	0.04	0	0.03	0.03	-0.23	-0.29	-0.29	-0.09	0.1	0.14	1

Table 4: Pearson Product-Moment Correlations for all questions in the ROLLING task.

N = Noticing; U = Understanding; WR = What-if reasoning; V = Identifying variables; G = Making generalizations; C = Imagining causes
(The numbers indicate the order of the question in the instrument; the order of the processes here corresponds to their order in the instrument).

APPENDIX G

MEAN AND STANDARD DEVIATIONS FOR LEARNING RESULTS.

Table 1: mean and standard deviations for those tasks which does not show statistically significant values.

TASK	Version	mean	s.d.
Balancing	I + G	5.0	5.2
	Str.	4.45	2.37
Black Boxes	I + G	8.16	6.07
	Str.	7.95	2.66
Floating	I + G	9.04	5.25
	Str.	7.41	3.1

APPENDIX H

TABLES FOR 'CONTROLLING CHANGES'.

		Change ø	Change 1	Change 2
Different configuration	After success	–	5(d) 4.2%	71 59.6%
	After failure	–	11 (d) 9.2%	32 26.8%
Same configuration	After success	1 0.4%	110 d = 4 w = 106 46.4%	16 6.7%
	After failure	8 3.3%	92 d = ø w = 92 38.8%	10 4.2%

Table 1: number of changes, and their respective percentages, introduced after having succeeded or failed when dealing with the same or different configurations in the BALANCING task (d = distance and w = weight).

		Change ø	Change 1	Change 2
Different configuration	After success	–	0.069 (5)	0.435 -0.145 (20)
	After failure	–	0.076 (6)	-0.272 (17)
Same configuration	After success	+	0.477 (23)	0.023 (8)
	After failure	0.044 (5)	0.398 (22)	-0.038 (5)

* Not taken into account for the purpose of being systematic in controlling variables.

Table 2: fractions' contribution to scoring in controlling variables in the BALANCING task (number of children actually performing on each situation are in brackets). Same and different configurations are taken separately. The fractions with minus sign mean lack of contribution in controlling variables.

	Change ø	Change 1	Change 2
Different configuration (box)	–	40 33.3% (16)	80 66.6% (23)
Same configuration (box)	315 31.9% (24)	671 68.0% (24)	–

Table 3: number of changes, and their respective percentages, introduced when dealing with the same or different configurations in the BLACK BOXES task. Number of children performing in each situation are in brackets.

		Change ø	Change 1	Change 2
Different configuration	After success	–	24 9.4%	43 16.9%
	After failure	–	75 29.6%	111 43.8%
Same configuration	After success	1 0.4%	6 2.6%	–
	After failure	2 0.8%	217 96.0%	–

Table 4: number of changes, and their respective percentages, introduced after having succeeded or failed with the same or different configurations (densities or solutions) in the FLOATING task.

		Change ø	Change 1	Change 2
Different configuration	After success	–	0.130 (12)	0.134 -0.134 (18)
	After failure	–	0.243 (14)	-0.353 (19)
Same configuration	After success	+	0.025 (4)	–
	After failure	-0.008 (2)	0.965 (23)	–

+ Not taken into account for the purpose of being systematic in controlling variables.

Table 5: fractions' contribution to scoring in controlling variables on the FLOATING task (actual number of children performing on each situation, are in brackets). Same and different configurations are given separately. The proportions with minus sign mean lack of contribution in controlling variables.

Different configuration	Change ø	Change 1	Change 2	Change 3	Change 4
After success	–	+ Gp = 22 Tv = 14 Th = 20 Ta = 7 63 24.6%	Gp,Tv = 12 Gp,Th = 5 Gp,Ta = 5 Tv,Th = 9 Tv,Ta = 5 Th,Ta = 8 44 17.1%	Gp,Tv,Th = 4 Gp,Tv,Ta = 4 Gp,Th,Ta = 3 Tv,Th,Ta = 2 13 5%	Gp,Tv,Th,Ta 2 0.7%
After failure	–	Gp = 46 Tv = 17 Th = 9 Ta = 9 81 31.6%	Gp,Tv = 11 Gp,Th = 8 Gp,Ta = 5 Tv,Th = 5 Tv,Ta = 4 Th,Ta = 7 40 15.6%	Gp,Tv,Th = 5 Gp,Tv,Ta = 5 Gp,Th,Ta = 1 Tv,Th,Ta = 1 12 4.6%	Gp,Tv,Th,Ta 1 0.4%

Same configuration	Change ø	Change 1	Change 2
After succes	*17 127 9.4%	D = 31 V = 159 *Ta = 3 *V = 1 194 14.3%	D,V 12 0.9%
After failure	*40 118 8.7%	D = 117 V = 706 *Ta = 7 *V = 2 832 61.6%	D,V *D & Ta = 2 68 5%

+ Gp = gun, position; Tv = trap, vertical; Th = trap, horizontal; Ta = trap, angle.
* Attempts made in impossible conditions due to the position of the trap.

Table 6: number of changes, and their respective percentages, introduced after having succeed or failed when dealing with the same or different configurations in the ROLLING task.

Different configuration

	Change ø	Change 1	Change 2	Change 3	Change 4
After success	–	0.294 (19)	0.069 -0.069 (15)	-0.050 (10)	-0.015 (2)
After failure	–	0.339 (21)	-0.128 (16)	-0.030 (6)	-0.001 (1)

Same configuration

	Change ø	Change 1	Change 2
After success	+	0.128 (18)	0.004 -0.004 (7)
After failure	-0.066 (16)	0.714 (24)	-0.078 (20)

+ Not taken into account for the purpose of being systematic in controlling variables.

Table 7: fractions' contribution to scoring in controlling variables on the ROLLING task (number of children actually performing in each situation are in brackets). Same and different configurations are given separately. The fractions with minus sign mean lack of contribution in controlling variables.

APPENDIX I

ILLUSTRATIONS OF THE TASKS.

Illustration 1: the BALANCING task.

Illustration 2: the BLACK BOXES task.

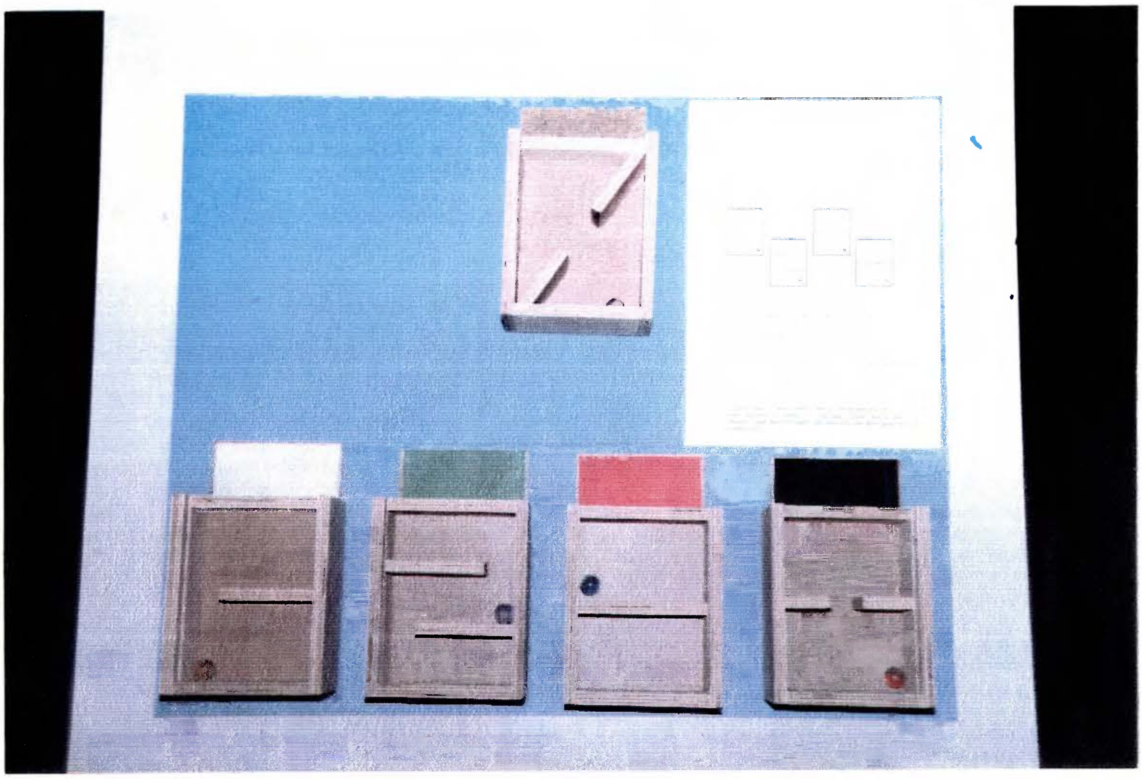


Illustration 3: the FLOATING task.

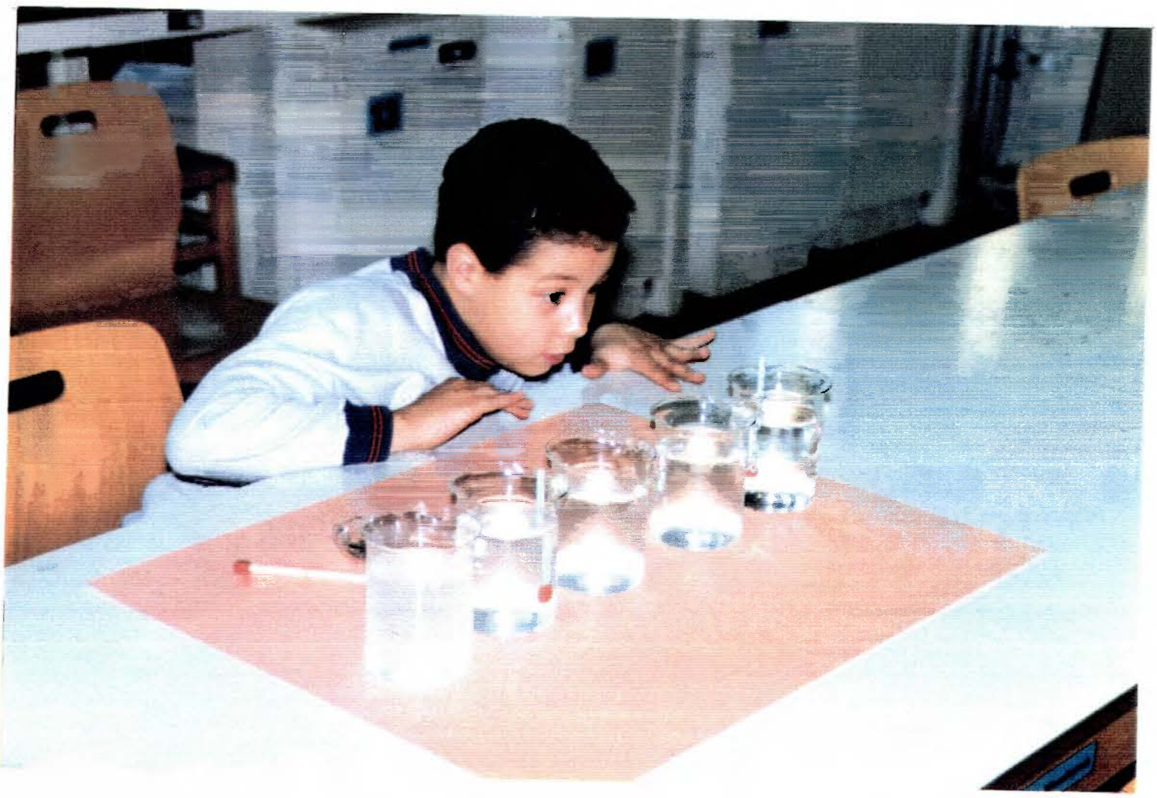


Illustration 4: the ROLLING task.