CHANGING TEACHER BEHAVIOUR AND PUPIL ATTAINMENT IN INQUIRY-BASED SOCIAL STUDIES LESSONS

A thesis
submitted in fulfilment of
the requirements for the Degree
of
Doctor of Philosophy
at the
University of Waikato
by
CLIVE FRANCIS McGEE

University of Waikato

1978

ABSTRACT

Over the last decade there has been a growing emphasis upon inquiry-based teaching in social studies. While there is no agreed definition of the inquiry method, there are particular goals and outcomes associated with this method of teaching and learning; fostering children's thinking processes is regarded as much more important than the accumulation of factual information. Asking questions has been recognised as one of the best ways to foster and develop inquiry learning in pupils, and yet previous research has shown that there is a need for teachers to ask a wider range of questions than they presently do.

One way to help teachers learn additional questioning skills is through an in-service programme. The present study was designed to find whether two kinds of in-service courses would modify teacher questioning. Twenty-four teachers in six Intermediate schools in a large New Zealand provincial city and a satellite town, were assigned to three treatment groups. The experimental treatments were: an in-service programme which taught specific questioning skills; another in-service course which dealt with inquiry teaching in a general way; and a no-treatment control group. The usual classes of the teachers formed the pupil sample.

Over a five-month period, each teacher conducted three 20-minute social studies discussion lessons which were taperecorded. Baseline data were collected in a pretreatment lesson, and posttreatment data were gathered from discussion lessons taken shortly after the experimental in-service courses and also after a 12-week delay.

As well as finding out whether in-service courses were followed by changes in teacher questioning, the experiment was also designed to see if pupil attainment was affected by any such changes. Parallel tests were constructed from the content of each lesson, and comprised four subtests which were intended to measure a range of thinking skills

considered important in inquiry-based social studies discussion.

From lesson transcripts, teacher questions were coded, using a classification system designed for this study. Statistical analyses indicated that the skills course teachers were able significantly to reduce their asking of cognitive memory and convergent questions, and to increase their use of evaluative and divergent kinds. They also talked much less, and encouraged more extended patterns of pupil discussion. In delayed lessons, the changes were maintained to a significant extent. The teachers who took the general course changed very little in comparison with the control group teachers.

Analysis of test scores indicated that the pupils of the teachers who took the skills course significantly increased their production of answers to free response items in the evaluative and divergent thinking subtests without decreasing other attainments.

The results of the experiment showed that in-service programmes which try to change teachers' questioning, and associated skills need to give attention to identifying the behaviours, communicating them clearly, providing practise opportunities, and giving feedback about performance. When teachers ask a wide range of questions in social studies lessons, pupils' evaluative and creative thinking can be increased, both in discussion and in free response tests.

ACKNOWLEDGEMENTS

I wish to express my indebtedness to my supervisors: Professor P.S. Freyberg for his continued counsel and valued assistance, and to Professor A.H. McNaughton for his appreciated critical comments.

I am most grateful to a number of colleagues for their assistance and advice about numerous aspects of this study, such as test construction and marking procedures.

I would like particularly to thank all those teachers, school principals, and pupils who so willingly cooperated.

I am appreciative, too, of the research leave granted by the Department of Education, Wellington, without which the collection of data would not have been possible, and to the New Zealand Educational Institute for research funds.

To my wife, Judy, I am grateful for her assistance with test administration and for her patience and support during the course of this study.

TABLE OF CONTENTS

VOLUME ONE

CHAPTER 1

THE PROBLEM

1.1 Introduction

Over the last 20 years or so there have been very substantial changes in the primary school curriculum and the methods used to teach it. Too often it appears that serving teachers are expected to 'pick up' the new ideas and skills required, but an increasing emphasis has lately been placed on in-service training to help teachers incorporate the innovations into their classroom behaviour. There has been little, if any, systematic evaluation, however, of the effectiveness of such in-service training programmes in New Zealand.

To take a specific instance of curriculum change and its consequences for teachers: recent developments in social studies have led to greater emphasis upon the teaching of social science concepts from several major disciplines. These concepts have become the basis for many new social studies instructional programmes in the United States (Thomas and Brubacher, 1971) and elsewhere.

Almost invariably, in these programmes, considerable emphasis is given to objectives that stress cognitive processes such as inquiry, problem-solving, creative and productive thinking, critical thinking, and discovery learning in the belief that social studies teaching should involve more than just the passing-on of information (Fenton, 1966; Jarolimek, 1969; Michaelis, 1975). For example, in New Zealand the *Draft Syllabus in Social Studies* (Department of Education, 1972, p.1) states:

> ... that as they make inquiries within a range of cultural settings, pupils will be gathering information. Using this information, and their own experiences, they will relate facts, and build up and express important ideas about human behaviour in their own terms. The forming and reshaping of ideas is a continuing process

> in which pupils will be using thinking skills as they move towards tenable conclusions and generalizations.

The stress upon inquiry processes is further shown by the importance given to the fostering of abilities that:

> enable pupils as they work with information to ask significant questions and become involved in problem solving.
>
> *(ibid.)*

These educational goals place considerable demands upon the teacher of social studies, particularly as they are associated with a new conceptual structure for content, and a reorganisation of content and learning which are probably unfamiliar to many teachers. It is now realised that exhortations to teachers to adopt innovations such as these, even when supported by written suggestions, are insufficient to produce marked changes in their classroom teaching practises (Wallen *et al.*, 1969). Thus if teachers are to learn about the new developments in a curriculum area — such as social studies — opportunities for in-service training have to be provided.

The effectiveness of in-service programmes for teachers has tended to be assumed rather than been demonstrated. Flanders (1970) observed that many in-service programmes are too general to effectively change teachers' classroom behaviour and it would appear that in New Zealand, also, many of the experiences have been of a general nature with little concentration upon helping teachers modify their teaching behaviours in the classroom itself. Associated with this concern about the generality of many programmes is a growing amount of research evidence which indicates that in-service courses which do focus upon specific teaching skills can result in substantial changes in teachers' classroom behaviour (for example, Taba, 1966; Borg, 1975).

Flanders (1970) considers that most teachers want to improve their own classroom effectiveness, and suggests that they would welcome specific in-service experiences which help them achieve this. In New Zealand there is some evidence that teachers attend in-service courses in the hope that they will learn specific teaching techniques. For instance, Courtney (1972) found that over half of a large sample (N = 187) of teachers attended in-service courses primarily to

'improve personal efficiency in the classroom'. Two-thirds of the classroom teachers rated the study of new teaching and learning methods which were directly related to their classrooms as their most vital concern.

One unanswered question, then, is: how can these expectations best be realised? In New Zealand, as elsewhere, there is a need to develop and try out a range of in-service models and approaches for this purpose. Timms (1975) listed three goals which might guide such programmes:

1. To upgrade and improve classroom instruction.
2. To act as a medium for the change process.
3. To provide the means to attain curricula changes.

1.2 Focus of the study

Accepting that 'the inquiry method' is now a recognised component of social studies teaching in New Zealand it would seem important that teachers be given in-service opportunities to learn to apply this method. And if what was said above about the need to focus upon specific skills is accepted, then in-service programmes would have to specify the teacher and pupil behaviours to be encouraged. It is widely recognised that teacher questioning represents a cluster of teacher behaviours essential to the implementation of inquiry teaching and learning (Ryan, 1971, Taba *et al.*, 1971). Furthermore, teacher questioning has been identified as an area which urgently needs in-service attention (Gall, 1970).

An appropriate vehicle, then, for the study of the effectiveness of an in-service programme in modifying the actual classroom practice of teachers would seem to be the specific questioning skills required to implement the inquiry method in social studies. Such an investigation should be able to indicate whether focussing upon specific teaching skills, such as questioning, is more effective than a general kind of course in helping teachers learn the techniques required to promote inquiry.

It would also seem opportune to attempt a more ambitious undertaking. After all, the prime purpose of teaching is to bring about some kind of pupil learning, and there would be no use in changing teacher behaviour if the changes did not in some beneficial way affect

pupil behaviour. As Peck and Tucker (1973, p.970) point out:

> It is simply not good science, nor good empirical practice, to recommend a way of training teachers without knowing, in fact, how it will affect those teachers' students. That, after all, is the only criterion by which the ultimate worth of any educational system can be measured.

However, as Rosenshine and Furst's (1971) review of the literature on teacher education has clearly indicated, little is known at present about the relationship between the classroom behaviour of teachers and pupil attainment. In addition, there is, in social studies, a dearth of tests which measure the wide range of children's thought processes that inquiry teaching emphasises (Wallen *et al.*, 1969).

It was decided, therefore, to try to devise an experiment which would not only analyse the extent to which teacher questioning and other associated verbal behaviours changed following different kinds of in-service courses, but also to discover whether teacher modifications to their verbal behaviours resulted in subsequent beneficial changes in pupil attainment.

1.3 Overview of the study

The design, implementation and results of the experimental study that was undertaken are reported in eight sections as outlined below.

Chapter 2 discusses the place of the inquiry method of learning and teaching in social studies and analyses the research literature on teacher questioning.

In Chapter 3 a review is presented of the available research on in-service education with particular reference to the impact of in-service programmes on teacher and pupil behaviour.

As a result of these literature reviews, a number of unanswered questions are identified in Chapter 4, and groups of research hypotheses relating to teacher and pupil variables which were tested experimentally in the present study, are stated.

To test these hypotheses an experimental design was devised based on a non-equivalent control group model, and this is outlined in Chapter 5, along with a description of the Form One and Two teachers and pupils who formed the sample, and an outline of the in-service

treatments given to two groups of teachers. The procedures used to collect taperecordings of classroom lessons, and the instruments employed to gather other teacher and pupil data are detailed in Chapter 6.

The results of statistical analyses of a number of null hypotheses relating to teacher and pupil behaviours are then presented in Chapter 7, and this is followed by a discussion of the results in Chapter 8.

In the final chapter the major conclusions to be drawn from the study are summarised, and a number of implications of the findings for in-service education and inquiry teaching in social studies are suggested.

CHAPTER 2

INQUIRY TEACHING AND TEACHER QUESTIONING

OVERVIEW: A review of the literature relating to inquiry teaching and teacher questioning is presented. It is shown that there is no accepted definition of the inquiry method, but there are identifiable elements relating to learning processes, teacher functions, and classroom climate. It is pointed out that in social studies there has been an increasing emphasis upon inquiry teaching. Teacher questioning is discussed as one of the most important skill clusters in promoting thinking processes in pupils. Purposes of teacher questioning are outlined, and types of questions that have been used in social studies and other research are given. Problems faced by researchers in this field are noted.

2.1 Inquiry teaching and learning

2.1.1 Inquiry learning as an educational goal

One of the major goals of education in general and of social studies in particular is the development of thoughtful inquiry in pupils which involves the development of critical thinking abilities, and in recent years there has been an emphasis upon discovery procedures in pupil-centred learning situations — or 'process orientated' programmes (Fair and Shaftel, 1967). This recent stress in social studies contrasts with what might be termed 'conventional' or 'traditional' approaches which have been 'content-oriented'. The latter type of teaching has placed primary importance upon the acquisition of knowledge, whereas the former stresses ways in which knowledge might be used effectively (Fenton, 1966; Massialas and Cox, 1966).

During the last decade, many published packages of social studies materials (and those in other subjects) for primary and secondary schools have been developed, which claim to reflect 'the inquiry method' of teaching (Sanders and Tanck, 1970). However, there is

confusion regarding the theory, objectives, and classroom practice of the inquiry method of teaching and learning. This is partly because there is no agreed operational definition for the terms used. For example, the term 'inquiry' has been used by many writers in a number of different senses. Kaufman (1972) has listed the following terms as having been used synonomously with 'inquiry method': inquiry, discovery, reflective thinking, inductive teaching, socratic teaching, problem-solving, inferential learning, and scientific inquiry.

Often in the literature no distinction is drawn between 'inquiry teaching' and 'inquiry learning' and 'the inquiry method'. In order to clarify some of this confusion, the inquiry method needs to be seen as part of the total teaching-learning situation, which can, in turn, be viewed as an interaction among teacher, pupil, and content. When planned effectively, the interaction is made up of three steps: developing objectives; planning and carrying out instruction; and measuring and evaluating results (Ober, Bentley and Miller, 1971).

Proponents of inquiry teaching and learning have varied in their precise interpretation of what the above steps should involve. Nevertheless, in philosophical and psychological theory related to 'the inquiry method' there are — in spite of definitional problems — certain broad elements which can be identified. These elements relate to the teacher, the pupil, and the classroom environment. It will also be seen, however, that the implementation of these elements varies from one instructional setting to another. Reference will therefore be made first to some early proponents of inquiry, and this will be followed by a review of more recent theory which has been influential in the emergence of the inquiry method.

2.1.2 The nature of inquiry teaching and learning

Educators have long recognised that the development of inquiry processes in pupils is an important educational goal. Socrates' teaching was motivated by a concern for producing a well-informed, capable Athenian. To encourage this development, verbal interchanges were organised between teacher and pupil, in which the teacher questioned skilfully, and the learner expressed ideas, tested and amended them, developed generalizations, and was led to improve his thinking skills. The student 'discovered' his own solution. So considerable stress was placed upon processes of thinking (Brubacher,

1963).

Francis Bacon attempted to instil the principles of logic through inquiry. His 'scientific logic' consisted of collecting data, eliminating unwanted data, and careful scrutinising of the remaining data. The method was aimed to produce what is possibly equivalent to an hypothesis — the 'first vintage' statement of the problem — from this scrutiny. Bacon extended the application of systematic logic from the analysis of propositions and discourse to the analysis of ways of knowing (Curtis and Boultwood, 1964). He was thus concerned with the structure of knowledge as well as with the process of learning.

Comenius also believed that teaching should lead to the development of a system of logic which could be applied to any body of knowledge: this would give the knowledge a structure and make it easier to learn. For Comenius, the teaching process itself should be logically structured, too, and his belief that teachers must liberate their pupils from fear (Curtis and Boultwood, 1964) is an important precursor to more recent theory which stresses a more open-minded learning atmosphere in the classroom.

The process of synthesising new knowledge from previously presented facts was seen by Herbart as important for learners. In this process, the learner formulates rules, laws and concepts, from previous observations. Herbart's belief in sequential learning was reflected in his attempt to develop a systematic, sequential method of teaching, in which the interests of pupils were of prime importance, and in which pupils were seen as individuals. Followers of Herbart have modified his original four-fold division of the process of synthesis or correlation into five stages, which could form a basis for inquiry. First, 'preparation' revised pupil knowledge to which new information would be related; second, 'presentation' gave pupils experience of new materials; third, 'association' enabled the pupils to analyse the new experience and compare and contrast it with the old; fourth, 'systematisation' revised the knowledge learnt; and in the fifth stage of 'application', the pupil used the new learnings or skill to apply to some situation (Curtis and Boultwood, 1964).

Of particular significance to inquiry teaching and learning was

Herbart's belief that experience is the basis of all education and development. He maintained that reasoning could follow only the development of an 'apperception mass' which is a complex structure of interrelated experiences or presentations. For the teacher using the Herbartian steps the greatest challenge was to organise experiences which were related to each other (Curtis and Boultwood, 1964). To organise in such a way is difficult, but the Herbartian stages represented a scientific approach to instruction in which the learner was intended to be an active inquirer.

Even in this cursory statement of the ideas of these early theorists, three important elements of inquiry can be identified: an emphasis upon finding ways to help the learner to structure his knowledge; the suggestion that problem-solving through a logical, sequential process is an effective way of developing such structures; and a stress upon using the interests and experiences of the child as a basis for inquiry.

One theorist who has had considerable impact upon the further development of the above ideas was Dewey. He provided (Dewey, 1915; 1916; 1933) a philosophical framework which has served as a stimulus for much of the subsequent work in inquiry and problem-solving techniques. He believed that it is essential for children to learn how to face up to problems; their ability to solve problems would help them become more effective citizens. Dewey (1916) saw too much emphasis — in schools — upon learning by rote memorization with the teacher acting in a restricted, authoritarian role as the imparter of knowledge. Dewey's learner-centred emphasis can be seen in his description of inquiry as

> the active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it and, the further conclusions to which it tends.
>
> (Dewey, 1933, p.9)

Dewey (1916) also viewed inquiry as an important natural tendency of children. This inquiry tendency, if linked with procedures for seeking answers, was seen as useful to the pupil in any situation that might confront him. Dewey therefore saw facts as the resources by which children could carry out new inquiries in order to lead to new information, concepts and generalizations. Facts were merely a means

to an end, and not an end in themselves.

For Dewey the most effective way to utilize children's natural tendency for inquiry was through problem-solving or 'reflective thinking'. He (Dewey, 1933, p.9) identified five stages in the problem-solving process:

1. Recognition of a problem.
2. Analysis of a problem.
3. Suggestions of possible solutions.
4. Testing of consequences.
5. Judgment of the selected solution.

Each of these stages requires the performance of particular cognitive tasks. During the first stage the problem is identified and spontaneous guesses are made which may or may not lead to a solution of the problem. Mental 'leaps' are often made. The second stage involves a more systematic and rational examination of the problem and the result is a more thorough definition of it. This grasp of the problem leads, in the third stage, to the formulation of working hypotheses which serve, in effect, as a research model to guide thinking progress towards a solution. In the fourth phase, the problem-solver attempts to reason out what might happen if each hypothesis is acted upon; and in the final phase the hypotheses are accepted, rejected, or modified as a result of non-acceptance. The thinking processes are therefore purposeful and involve several distinguishable cognitive tasks such as the use of reason, inference and generalization.

Dewey pointed out that the problem-solving process described above does not necessarily operate in the strict sequence that five stages might suggest. Nevertheless, he argued that no matter how the order is perceived, the intellectual traits he identified are necessary for reflective thinking to occur.

Dewey's problem-solving procedure is therefore less rigidly-conceived than the Herbartian steps cited earlier. Dewey stressed the need for the teacher to be flexible, whereas the teacher using Herbart's steps was still largely an imparter of knowledge which led to preconceived solutions.

In selecting problems for children to grapple with, Dewey (1933)

believed that the child's whole experience should be the source of the problems. In this respect Dewey has had an influence on inter-disciplinary approaches to organising content, which contrasts with Bruner's (1960; 1966) later emphasis upon utilizing the structure of certain academic disciplines as the source of problems. This will be discussed later in this section.

Another American who attempted to develop Dewey's ideas on problem-solving and inquiry in the classroom was Kilpatrick. He was, like Dewey, critical of much of the teaching of his time because of what he saw as its overriding concentration upon imparting knowledge. He argued that the main emphasis in teaching should be upon developing a 'total personality' by getting children to think more deeply about their own world, utilizing creativity, persistence, and sensitivity to 'straighten out some intellectual difficulty' (Broudy and Palmer, 1965). For Kilpatrick, Dewey's problem-solving method became one 'project method'. The teacher's task was to help the child define the problem so that the child could then try independently to reach a solution. What was often termed 'the activity classroom' was viewed by Kilpatrick as a place of pupil discussion, planning and overt movement, and less as a place of teacher telling and pupil reciting.

Bruner (1961) has had considerable influence upon recent curriculum developments with his advocacy of discovery learning as the most effective means of developing knowledge structures. In effect, this means that use is made of creativity and problem-solving, and children are encouraged to learn how to inquire into problems in an autonomous way. Four advantages which are claimed by Bruner for the discovery approach have relevance for the inquiry method of teaching and learning. These are an increased facilitation of transfer of learning; the acquisition of problem-solving skills; an increase in intrinsic motivation; and, increasing efficiency in memorising what has been learnt. These advantages will be discussed further in Section 2.1.4.

From a rather different standpoint, Guilford's (1959) structure of intellect model is a comprehensive description of cognitive processes which has proved to be important in the development of inquiry teaching techniques by distinguishing between convergent and divergent thinking. Convergent thinking implies a search for the

right or known answers, whereas divergent thinking suggests the generation of new ideas using abilities such as flexibility, fluency, originality, and elaboration of ideas.

Researchers (for example, Gallagher, 1965) who have utilized Guilford's model in analysing discussion lessons in a variety of subjects have concluded that there is a need for greater use of creativity skills such as divergent thinking, and Guilford (1959) himself advocated greater use of creativity in classrooms. He outlined four steps that he believed were essential to the creative act of problem-solving. These steps can be summarised as follows:

1. A period of preparation in which the problem is considered.
2. A period of incubation in which there is little apparent progress made towards a solution.
3. An inspiration which produces a partial or final solution to the problem.
4. An evaluation or verification period, in which the problem-solver tests the solution and tries to determine the value or appropriateness of the solution.

Guilford's steps are similar to those of Dewey's problem-solving phases. The existence, in most cases, of an incubation stage had been suggested by Wallas in 1926 (reported in Hallman, 1967); he saw it as consisting of a series of spontaneous, uncontrollable events or explorations, as precursors to reaching solutions. Rogers (1959) called this phase 'openness to experience'; a time when a person is receptive to new ideas. Hallman (1967) also emphasised that such openness is essential for the development of creativity.

According to Massialas and Zevin (1967, p.6) inquiry is 'behaviour which is characterised by a careful exploration of alternatives in seeking solutions to a problem'. Not only did they see problem-solving as a key process of the inquiry method, but they placed stress upon the consideration of alternatives, conjectural thinking, and the use of creative ideas which suggests a requirement of flexibility of thought. Shulman (in Ryan, 1969, p.20) also stressed both the problem-solving aspect and creative thinking when defining inquiry as 'the process of coming to grips with problematic situations

which require the discovery of available techniques or the invention of new means for their resolution'. In Shulman's definition, then, there is a suggestion that original thinking is also necessary. Simon (1967) also hypothesised that the process a person uses in carrying out creative thinking is indistinguishable from ordinary problem-solving processes. He said that what might distinguish the two is the product rather than the process — that the creative thinker produces novel solutions.

It has been argued, then, that there is a close connection between problem-solving and creativity. This relationship can be further seen in Rossman's (1931) postulated seven stages of creative production:

1. Observation of the need and difficulty.
2. Analysis of the need.
3. Survey of available information.
4. Formulation of objective solutions.
5. Critical analysis of the solutions.
6. The birth of a new invention — the idea proper.
7. Experimentation to test the idea.

Again, these stages are similar to Dewey's stages in that they stress a method of investigation. However, it should be pointed out that although creative elements are involved in problem-solving, not all creativity need be connected with solving identifiable problems (Torrance, 1962).

It is apparent from what has been said above that a number of theorists see value in children discovering things for themselves, rather than being 'instructed'. Taking a flexible view, Gagné (1966) argued that discovery learning processes may occur as a part of most kinds of learning acts, whether simple or complex. The process of discovery relates, he said, to events which are internal to the learner, and the events involve a search for, and a selection of, data. Both processes take place within the individual learner, therefore discovery learning is idiosyncratic. Suchman (1962) also stressed that creative thinking was autonomous, or self-directed, and that creative thinking produces new forms of awareness for the learner.

This places a stress upon the objective of improving the process

of thinking in the belief that the acquisition of information will also occur in inquiry lessons: 'The educator should be concerned above all with the child's process of thinking, trusting that the growth of knowledge will follow in the wake of inquiry' (Suchman, 1961, p.151). This reflects Suchman's confidence that knowledge growth can be promoted by emphasis upon learner inquiry.

Postman and Weingartner (1969) emphasised that the focus of intellectual energy should be upon the active investigation of structures and relationships rather than upon passive reception of other people's ideas. This stress is, therefore, upon the pupil generating his own ideas, and in doing so, learning how to learn.

From what has been said so far, it can be seen that a major goal for inquiry teaching is to foster a particular kind of learning in pupils by getting them to use problem solving, creative and discovery methods. However, these processes also represent outcomes of inquiry. For example, Bruner (1963, p.22) defined the result of discovery learning as 'a matter of rearranging or transforming evidence in such a way that one is enabled to go beyond the evidence so reassembled to additional new insights'.

An important consideration for teachers who use inquiry teaching is to know when and how to teach ideas and concepts. Bruner (1960, p.33) claimed that 'any idea can be represented honestly and usefully in the thought forms of children of school age ...'. He argued that because children progress from motoric or sensory (enative) representation to representation in the form of relatively concrete images (iconic) to abstract representation (symbolic), it follows that the sequence in teaching should be the same. Thus, for Bruner, subject matter can be broken down into a progression from simple to complex concepts. Additionally, he argues (Bruner, 1966) that the three-stage pattern above may be used for sequencing individual instructional units for those pupils who primarily employ symbolic modes of learning.

Although Piaget himself has not wished to explore its educational implications, his work, too, has particular significance for inquiry teaching and learning in deciding when ideas and concepts can be introduced to pupils. Piaget (1950) has described a fixed developmental sequence in children's conceptual constructions. The first stage is the sensory-motor (from birth to 2 years of age) in which

the coordination of the various senses is accomplished and a relationship established between perception and action. The pre-operational stage (2-7 years of age) is the period of acquiring a language which permits the child to interact with the world symbolically instead of directly through motor action. The third stage (7-11 years of age) is characterised by a developing dependence on perceptions and a growing capacity to reason, although at a concrete level only.

In the fourth stage — formal operations (11-15 years of age) — the child deals with what might be whereas in the previous stage he dealt with what is (McGuire and Rowland, 1971). The significance of this for inquiry-based lessons is that the child attempts problem solving by making abstract considerations of possible solutions, according to the relationships between the elements of the problem.

By developing hypothetico-deductive reasoning abilities in the formal operations stage the learner is able to formulate basic postulates from which he deduces testable hypotheses. After testing an hypothesis the learner either accepts it as a solution to the problem or goes back to modify the approach taken, thus once again initiating the hypothetico-deductive processes. Lefrancois (1975), amongst many others, has cautioned, however, that it should be remembered that not all children pass through Piaget's stages at the same ages.

An important question in arranging instructional procedures is whether or not children's thinking can be accelerated so that the formal operations stage, for example, is reached earlier than might otherwise be expected. It seems — by implication — that a rich experiential background should lead to the earlier appearance of thought processes of a higher stage. In this respect Piaget's notion of fixed developmental stages is not consistent with Bruner's (1960) claim that any idea or subject can be taught to any child at any stage of development. Bruner contended that the intellectual development of young children could be speeded up by the introduction of a variety of instructional materials. The exploration of these and the subsequent discovery of underlying principles would move him into the higher stage of development. As Lefrancois (1975) has pointed out, however, there is inconsistency in research studies so that no conclusive evidence is yet available, one way or the other.

Gagné (1966), on the other hand, has maintained that human development results from cumulative learning. Advanced cognitive functioning depends upon a large number of less advanced abilities which the child acquires sequentially as he develops. The entire learning sequence has been described by Gagné in terms of eight distinct types of learning from stimulus response at the lowest level to problem-solving at the highest level of complexity.

From combinations of simple principles more complex principles are formed. The 'stage' of a learner depends less upon his age (as tends to be the case with Piaget's stages), than upon the relevant capabilities he has at that time and how much he has to learn before a new goal is achieved. The patterns or structures generated in this way build upon each other and transfer learning occurs among them, so that an ever-increasing intellectual competence is generated. Gagné's theory contrasts with Piaget's more rigid stages of cognitive growth, and unlike Piaget, Gagné does not postulate a strict order of appearance of structures in the learner.

Taba (1966) considered that thinking can be taught and (like Piaget) that the processes of thought develop sequentially so that each successive step is dependent upon the mastery of the preceding one. She saw this process in terms of three basic cognitive skills: concept formation, which involves the use of mental skills such as enumerating, labelling and ordering; generalizing and inferring through the interpretation of raw data; and the application of principles and facts to explain and predict new phenomena.

These cognitive skills, translated into teaching tasks, have been utilized in the field of social studies (Taba *et al.*, 1971). Carefully sequenced content was formulated and taught inductively. The teacher's role was seen as primarily that of an asker of questions which Taba viewed as central to the inquiry process.

It is evident, too, that Taba placed considerable importance upon children learning ideas and concepts that had been carefully selected from their respective disciplines and it was noted earlier that theorists have long been concerned with how best to structure ideas and concepts so that they can be learnt most effectively. Bruner (1960, p.33) has argued that 'the task of teaching a subject to a child at any particular age is one of representing the structure of that subject in

terms of the child's way of viewing things'. By structure Bruner meant that different academic disciplines are founded on particular ideas about how subject matter is related. For example, in mathematics Bruner (1961, p.7) stated that 'algebra is a way of arranging knowns and unknowns in equations so that the unknowns are made knowable'. If a pupil grasps the fundamental mathematical ideas in an equation he is able to work out new equations. That is, a pupil is able to transfer what he has learnt in one situation and apply that knowledge in another setting to solve another problem.

Bruner therefore believed that the pupil would use the structure he had learned to create new structures — not by being taught to memorise them, but by discovering or re-discovering them for himself. Bruner (1960, p.33) summarised as follows: 'To learn structure, in short, is to learn how things are related'.

Unlike Bruner, Kaplan (1964) did not see that structure was so important, stating that truth has no boundaries and that each discipline shares with others the concepts, laws, data, models, theories and explanations that contribute to inquiry. Kaplan has emphasised the wide use of imagination and intuition, in contrast to Bruner's stress upon the child's recognition of structure. And although Dewey considered that knowledge is structured, he did not believe that the structure could be imposed from outside the learner. Instead, he believed that the structure is developed by the learner himself.

> The child is the starting point, the center and the end. His development, his growth, is the ideal! ... self-realisation is the goal ... learning is active. It involves reaching out of the mind. It is he (the child) and not the subject-matter which determines both quality and quantity of learning.
>
> (Dewey, 1915, p.9)

2.1.3 Strategies of inquiry teaching and learning

Thus the current emphasis upon inquiry teaching and learning in a number of subject areas has developed out of the ideas of a number of educational and psychological theorists such as those cited above. Although, as Kaufman (1972) has pointed out, the various theorists' contributions have not been synthesised into an adequate model for describing the theory of inquiry teaching, there are, nevertheless, again certain elements which can be identified. These relate to the

teacher, the learner, and the classroom environment.

First of all, the lack of a single model is partly explained by the fact that the conceptions of inquiry almost necessarily vary in different situations, such as subject area, level of class, and stage of instruction. To search for 'the ideal' method of inquiry might therefore be futile at the present stage of what is known about the teaching-learning situation. Perhaps a more fruitful approach would be to explore inquiry teaching and learning in particular settings. Three particular approaches will thus be briefly described: the MACOS programme (Bruner, 1966); 'analytical', 'discovery', and 'social problems' episodes (Massialas and Zevin, 1967); and inquiry training (Suchman, 1961).

One example of an inquiry programme in social studies is *Man: A Course of Study* (*MACOS*), developed through the initiative of Bruner (1966) for Grade 5 pupils. The programme emphasised inquiry teaching strategies by giving pupils opportunities to work with supplied data to make and evaluate educated guesses through discussion. The unifying themes of the programme are the basic questions: What is human about human beings? How did they get that way? How can they be made more so? (*ibid.*, p.74). Five topics were developed to explore the three questions just cited: tool making, language, social organisation, the management of man's prolonged childhood, and man's urge to explain his world. The first part of the teaching involved the pupils grasping the fundamental concepts relevant to these topics, followed by participation in 'informed guessing' and hypothesis making, and finally, stimulating pupil 'self-consciousness' about thinking which involves a degree of tentativeness and competence at finding out for themselves. Once again, there is therefore considerable stress upon the process of learning and less emphasis upon teaching pupils to remember isolated facts. The MACOS programme is thus an attempt to translate notions about how children inquire into particular instructional sequences that might promote the inquiry process.

In a study reported by Massialas and Zevin (1967) discussions with high school students covering a number of subjects were conducted by Zevin. The discussions were of three kinds, all of which were claimed by the authors to involve inquiry.

Analytical episodes involved discussions in which the teacher

tried to take the students through a series of sequential and orderly steps from the identification of a problem to the finding of a solution, for example, selecting the most likely place for human settlement on an unknown map.

Discovery episodes gave the pupil opportunities to 'make leaps into the unknown and uncontrolled world, and he learns the value of formulating plausible hypotheses about human interactions' (*ibid.*, p.73). Thus these episodes were designed to involve pupils in what Bruner (1961) termed 'figuring out'. For example, in one episode the pupils were shown a statuette which was strange to them and asked to find out what it was. The teacher's role was to help provide a classroom climate conducive to the open-ended inquiry which followed.

The third set of episodes involved the examination — through discussion — of social problems in which pupils were encouraged to give their personal value judgments.

Suchman (1961) developed a particular technique of inquiry amongst Intermediate level pupils which varied yet again from those given above. In physics lessons pupils were given a specific problem involving causality and asked to discover why the demonstration had the results that it did. To move towards a solution the pupils had to gain additional information by asking questions to which the teacher only answered 'yes' or 'no'.

Suchman claimed that he was developing 'autonomous inquirers' and although many children were at first ill-at-ease, he believed that through hypothesising, the children were performing formal logical thinking operations.

In the above examples of inquiry teaching and learning it can be seen that there are variations in the precise nature of the respective roles of the teacher and the pupils. For example, in Suchman's technique, the child has the responsibility for asking questions whereas in most of the other approaches the teacher has the major responsibility. The broader elements however show greater similarity. The learner is regarded as an active enquirer, the teacher as the person responsible for making the inquiry possible by setting-up the necessary conditions, and the classroom climate is one in which pupils feel free to participate fully in open-ended inquiry.

2.1.4 Inquiry: its effectiveness

Although there have been many proponents of inquiry teaching, its direct theoretical and empirical justification has been rather thin. Usually the merits of inquiry over other methods of teaching have been argued by implication alone.

Bruner (1961) has put forward a number of benefits which he believed accrue from discovery (or inquiry) teaching and learning. He indicated that when pupils discover things for themselves they increase their intellectual 'potency' by being able to develop problem-solving search patterns. This allows the learner to transform and organise information.

A second advantage cited by Bruner is that pupils learn 'the heuristics of discovery' which means that the pupil can best learn how to discover new information if he has had practice with effective discovery procedures.

Third, Bruner argued that by discovering autonomously pupils would shift from being extrinsically motivated to being intrinsically motivated. He pointed out that pupils who are accustomed to approval of parents and teachers for 'the right way to do it' may develop conformity patterns so that they avoid failure and do not take risks in their thinking. Under the no-risk circumstances children would not recognise the intrinsic rewards that inquiry and finding out for oneself bring.

Kagan (1966) too, had been aware of this tendency and claimed that as early as kindergarten years, children develop different degrees of success or failure expectancy. He noted that a challenge for classroom teachers is to minimize the extent of humiliation that accompanies failure. Excessive competitiveness too, was seen by Kagan as an inhibiting factor to discovery and inquiry learning, because it tends to accentuate the pupils' search for correct answers when teachers reward only 'correct answer' behaviour.

It has been claimed that the arousal of children's curiosity is a means of making learning more effective (for example, Dewey, 1933; Guilford, 1959; Massialas and Zevin, 1967).

Day and Berlyne (1971) have stated that a teacher has two ways of arousing pupils: either by using a system of rewards and punishments

meted out by himself, or by manipulating the novelty and complexity of lesson material to maximise arousal. In the latter situation, no threats are used, and extrinsic motivation is replaced by intrinsic motivation. Like Bruner, Day and Berlyne believed that every person has some degree of creativity, and therefore they argued that intrinsic motivation is a reachable goal for all pupils. And yet it was pointed out that curious children do not necessarily show up as high achievers on traditional classroom measures such as achievement or intelligence tests; nor are they always the ones who are conspicuous in lessons. It does seem, however, that they are intrinsically motivated (*ibid.*, p.322). The classroom neglect of curiosity was viewed seriously: 'One must alarmingly conclude that though we treat creativity as a goal of education, by failing to reinforce curiosity, we fail to reinforce creativity' (*ibid.*, p.332).

Torrance (1967), too, accepted that all children possess some degree of creativity, and that it can be actively developed by teachers, and he claimed that in American schools, at least, the level of creative thinking falls as pupils grow older, because it is not fostered. Yet, as Parnes (1967) pointed out, although we still know little about what 'creativity' really is, we <u>do</u> know how to stimulate greater creative behaviour in people.

Some empirical evidence that motivation was improved as a result of inquiry-type lessons was noted also by Suchman (1962). In an experiment with Grade 6 pupils, Suchman found that those who had worked on the Inquiry Training method were more actively involved in their studies than other children, maintained a high level of interest and excitement during the 25-week period of training, became more fluent in their production of questions, and were more analytical in probing ideas. He proposed that children should be given a variety of experiences to use creative thinking as part of the process of inquiry learning, so that they could find out for themselves what works best in each situation met.

Critics of teaching approaches which rely heavily on inquiry learning and discovery have claimed that too much time-wasting is involved and that pupils who are always forced to find out for themsevles may not obtain essential learnings. Friedlander (1965) argued that the 'trial and error' nature of discovery may produce confusion in pupils' minds. Also, pupils may not possess the powers of judgment

and evaluation that are necessary to assess their own intuitive reasoning.

Friedlander also claimed that there is no hard evidence that what a child learns from his own attempts at discovery are remembered any better than when learnt by other methods. Bruner (1961), on the other hand, argued that by discovering for themselves children embed new information in their personal cognitive structures and so improve their ability to retain and recall them, but he cites no evidence to back up his claim.

Associated with this supposed advantage Bruner (1957) has further argued that discovery allows the learner 'to go beyond the information given and make educated guesses' on the basis of remembered information. In social studies this would appear to be of particular significance, for many of the problems concerning human cultures and behaviour that are studied do not have absolute answers.

After weighing up respective viewpoints about the merits of discovery and inquiry learning, Shulman and Keislar (1966) concluded that the matter is most frequently decided on rational or emotional grounds rather than on findings from empirical research. And having reviewed a large number of recent investigations into inquiry methods in social studies Marsh (1975) came to the conclusion that although there is some evidence that inquiry is more effective than other methods in terms of 'recall, transfer and retention of data and ... developing specific skills in questioning and concept building' (p.281) the evidence was by no means conclusive.

2.1.5 Summary

In this section it has been pointed out that the major goal of inquiry teaching and learning is the development of higher levels of thought in pupils by involving them in problem-solving experiences. However, there is no widely accepted definition of exactly what constitutes 'the method'. An examination of relevant philosophical and psychological theory shows that in spite of this definitional problem there are certain identifiable elements in inquiry teaching. First, subject matter is organised or structured in ways that allow teachers to introduce concepts to pupils at appropriate times. Second, it is claimed that content and concepts can be very effectively taught and learnt by the use of problem solving, creativity and

discovery learning. Third, the teachers' function in inquiry lessons is to provide a flexible, open-ended classroom environment and engage children in the inquiry processes mentioned. Finally, the inquiry method stresses the processes of learning to a greater extent than the acquisition of knowledge alone.

Early exemplars of inquiry (for example, Bacon; Comenius; Herbart) stressed the use of logic as the basis for reaching solutions to problems. Herbart's teaching steps reflected the belief that both mental processes and content were governed by laws of logic. Dewey provided a major impetus for later developments in inquiry by demonstrating that it involves the resolution of perplexing problems by the use of logical, scientific investigation which capitalizes upon the child's natural curiosity. Kilpatrick's 'project method' is an example of the implementation of Dewey's problem-solving method, but his stress was more on flexible 'activity' methods. More recently Guiford's theory of intellect has provided a further stimulus for inquiry through the identification of a number of thinking operations considered suitable for inquiry learning and creativity, particularly the use of divergent thinking.

In the literature there are different viewpoints about how concepts can best be taught and learnt. Bruner has argued that each subject can be broken down into conceptual forms that can be learnt by children through discovery. Piaget has talked in terms of fixed stages of development which place limitations upon when concepts can be learnt. Taba, on the other hand, has tried to demonstrate that the learning of concepts can be accelerated by carefully structuring and sequencing instruction.

Using ideas such as these, various researchers have devised different versions of 'inquiry' (for example, Suchman, 1961; Bruner, 1966; Massialas and Zevin, 1967). These programmes show that while the underlying elements of inquiry outlined above are similar in each case, the actual implementation varies in relation to the specific objectives of the programme. For example, in one setting described by Massialas and Zevin the children tried to solve a problem by making 'hunches'; in another by being led through a series of questions by the teacher.

Proponents of inquiry methods have claimed that they have a number of benefits for learners. Bruner (1961), for example, argued that inquiry teaching improves problem-solving abilities, motivation, recall and the techniques of discovery and Kagan (1966) claimed that inquiry teaching reduces failure expectancy in pupils. However, little empirical evidence is at present available to sustain these assertions.

From what has been said it is clear that the teacher's role in inquiry lessons is crucial. In the following section this aspect is examined, and particular reference made to social studies classrooms.

2.2 Inquiry and the classroom teacher

2.2.1 General conditions for inquiry teaching

The primary function of teaching is to facilitate learning, and as in all instruction, the teacher's role is of fundamental importance in inquiry-based lessons. In order for inquiry learning to take place in the classroom, the teacher's responsibility involves the promotion of a suitable climate and the fostering of certain cognitive skills. One writer believed that these goals could be met by a teacher who:

1. Utilizes higher-level questions.
2. Solicits multiple and varied ideas from students.
3. Is receptive and supportive of all ideas expressed.
4. Is careful not to sway any student opinion through expression of self-opinion.
5. Manifests a respect for everyone's ideas even when they are in conflict with his own.
6. Encourages students to examine the reasons behind their responses.
7. Encourages students to listen and react to the ideas of others.
8. Provides opportunities for students to generate and and express their own ideas, rather than merely repeating the ideas of others.
9. Allows students to disagree with the ideas of others, including those of the teacher and those expressed by any media.

(Ryan, 1971, p.136)

These criteria are particularly significant because they indicate that many teachers need to change not only their cognitive behaviours in

relation to their pupils, but also their attitudes. 'Attitudes of teachers are the most important characteristic of the inquiry environment' (Postman and Weingartner, 1969, p.41).

Suchman (1961) also recognised that both attitudinal and cognitive skill changes were necessary for teachers to utilize the inquiry method. A teacher must 'abandon his traditionally directive mode and structure an environment that is responsive to the child's quests for information' (p.151). He must also recognise that success is important for the child. Through his experimental work in physics, Suchman believed children should be taught a broad schema of inquiry which helps them identify a sequence of goals toward which inquiry should be developed for maximum productivity.

Facilitation of inquiry appears to have particular significance for social studies education. Whereas the physics students are looking for physical causality, social studies pupils seek indications of social causality, through the identification of social situations and problems and discussion of data related to these. While their teachers hope that physics students will discover universal physical principles, they hope that social studies pupils will discover ideas and concepts in relation to such themes as those outlined in the Taba Curriculum for elementary grades (Taba *et al.*, 1971). Because social studies concepts are more diffuse they require more extensive exploration.

Another essential element in the teacher's role in inquiry teaching is that of an arouser of pupils' curiosity and interest to the point where pupils themselves start to identify problems or raise questions. The importance of curiosity in inquiry has already been stressed. To successfully arouse pupils' curiosity, a teacher is committed to a continual search for a more comprehensive understanding of the interests of each pupil. This understanding is more likely to lead to one of the major goals of inquiry teaching and learning: 'providing pupils with an organised, improved method of thinking about and dealing with information' (Skeel and Decaroli, 1969, p.548).

The importance of teacher questioning skills has been stressed by Postman and Weingartner (1969, p.41): '[The teacher's] basic mode of discourse with his students is questioning'. This implies that teachers need to use a range of question-types, for Postman and Weingartner add: 'Generally, he does not accept a simple statement as

an answer to a question' (*ibid.*, p.17). Teacher talk, therefore, should not dominate lessons which claim to teach by inquiry methods. '[The teacher] is engaged in exploring the way students think; not what they should think ... that is why he spends more of his time listening to students than talking to or at them' (*ibid.*). Massialas, Sprague and Hurst (1975) also have pointed to the need for teachers to reduce their amount of talk as they and their pupils jointly explore problems in a psychologically open climate.

For teachers unused to developing inquiry learning in their pupils, the adoption of such techniques may be difficult. For example, Lawton *et al.* (1971, p.16) pointed out that it is difficult for a traditional teacher to develop an inquiry approach which is centred on problem solving, but added that without such an approach 'the value of social studies teaching is dubious'. They stress again the importance of appropriate questioning by teachers, together with the need for the teachers to help their pupils to develop habits of tentativeness, so that information is always subject to modification in the light of new evidence. They point out that teaching skills required are probably difficult to learn, and teachers would need help in doing so.

2.2.2 Inquiry and social studies teaching

Most of what has been said thus far in this chapter about inquiry is applicable to social studies. As in other subject areas, there is no agreed definition of what inquiry in social studies is. Massialas and Sprague (1974) and Marsh (1975) saw the inquiry method as having two dimensions: procedural, and content or substantive. Both dimensions are important because they dictate the way in which a teacher conducts inquiry lessons; how the objectives for a lesson are set, and how the learning experiences are organised to realise the objectives. The procedural dimension requires students to undertake a sequence of activities that are designed to lead to the solution of some felt need or problem. The content dimension of inquiry relates to the subject discipline from which the problem under study has been drawn. As Marsh (1975) indicated, a wide range of structuring of learning procedures in social studies is possible, from highly structured to very open-ended situations, and the range of problems suitable for inquiry teaching is almost limitless. On these grounds, Marsh argued that there are good reasons for accepting a wide

definition of inquiry.

The components identified above indicate that inquiry teaching requires a set of conditions which include, first, a teacher who is able to select appropriate content and objectives, to organise suitable learning experiences, and to establish a flexible classroom atmosphere. Second, pupils need to engage in a range of thinking processes.

Numerous writers in social studies education have stressed the need for an emphasis upon the development of thought processes in pupils as a means of increasing individual's decision-making power, and to foster autonomous thinking (Benne, 1967; Estvan, 1968, Beyer, 1971; Ploghoft and Shuster, 1971; Ryan, 1071; Taba *et al.*, 1971; Traill *et al.*, 1972). Michaelis (1972), for example, has outlined the following thinking processes as being essential to the inquiry learner: recalling, observing, comparing and contrasting, classifying, defining, interpreting, generalizing, inferring, hypothesising, predicting, analysing, synthesising, and evaluating. The challenge for teachers, Michaelis argued, is to present pupils in social studies with a rich variety of experiences which would expose them to many new concepts, for concepts form the basis of inquiry thinking. Therefore 'the inquiry-oriented teacher is as conscious of the process of acquiring knowledge as he is of the knowledge itself' (Rogers and Genovese, 1969, p.538).

Analysing the views of these writers, and a review of inquiry literature by Kaufman (1972), suggested that six cognitive objectives are commonly stressed:

1. Less time should be spent on the recalling of facts in social studies lessons.
2. Pupils should be permitted to make decisions about which knowledge is relevant for them.
3. Inquiry should promote transfer of social studies knowledge.
4. Critical thinking by pupils should be a basic goal of the inquiry method.
5. Creative thinking and discovery learning are part of the inquiry process.
6. Pupils should be given the chance to use evaluation skills

by making choices and judgments.

In order to attain such objectives, the teacher has to translate them into learning activities. One way of doing this in social studies was suggested by Carpenter (1967) who nominated four stages in children's inquiry learning. They are summarised as follows:

1. Awareness and identification — a feeling of disturbance, perception of the elements involved, and clarification of the problem.
2. Data processing — collecting, organising, and evaluating relevant information.
3. Analytical synthesising — comparing, contrasting, inferring, speculating, and applying data in hypothesis formation.
4. Critical testing — probing the validity of hypotheses, reconstructing data, and reaching conclusions, which are often tentative and can be used in solving further problems.

Carpenter warns that the stages do not necessarily proceed neatly in the order stated but they can form the basis for lesson sequences.

In another attempt to sum up how social studies inquiry learning might operate, Massialas and Sprague (1974) identified the following classroom conditions which they considered necessary for discussions of controversial issues. First, the classroom climate should be psychologically open, and discussion should have a sense of purpose for the pupils. Second, the discussion issue should be presented by the teacher in a problem context as a springboard to further study. Third, the major instructional goal should be to get pupils to clarify some issue, to hypothesise about solutions, and to seek solutions. Fourth, the teachers' role should be to create the conditions from which problem solving may develop and to make available the resources to research from. Question-asking is an essential skill in this role. The final condition is that pupils should report back about their findings after having gathered information, and the teacher should then adopt a suggesting role.

In order to bring about such a set of conditions, learners need to be fully involved in the inquiry process. There is no simple set of criteria to describe what a 'good' or ideal inquirer might be. However, the following characteristics of 'good' inquiry learners

reflect the above components, and are summarised from Postman and Weingartner (1969, p.41). Good learners are children who:

1. have confidence;
2. enjoy solving problems;
3. seem to know what is relevant to their survival and what is not;
4. prefer to rely on their own judgment;
5. usually are not afraid of being wrong;
6. are not fast answerers;
7. are flexible;
8. do not have to have a final solution to every problem; and
9. have a high degree of respect for facts, and can distinguish between fact and opinion.

These authors have stressed the need for a flexible, relaxed class atmosphere in which children feel confident about their own ability to contribute, without necessarily always being 'right' in their answers or ideas.

To summarise the rather scant literature on the direct application of inquiry methods to social studies teaching, then, it emphasises a teacher role which:

(a) sets up (or helps set up) problems for study;
(b) leads to joint exploration of each problem with his pupils;
(c) develops his pupils' thinking processes; and
(d) involves teaching in an open-ended atmosphere in which the children feel free to participate.

Furthermore, as has also been pointed out, teachers may require help in changing their teaching behaviours to make this kind of inquiry teaching possible. For example, questioning involves an important set of skills associated with developing thinking processes, and yet, as will be seen in the next section, teachers characteristically ask a narrow range of questions, even though they have frequently been exhorted to do otherwise. To help teachers adopt inquiry methods it may thus be necessary first to help them improve their questioning techniques.

2.3 Teacher questioning

Numerous theorists referred to above (for example, Bruner, 1966; Taba, 1966, Massialas and Zevin, 1967; Postman and Weingartner, 1969) view the asking of questions by teachers as central to inquiry teaching and learning. Indeed, Taba *et al.* (1964) asserted that the teacher's technique of asking questions was by far the most influential single factor in inquiry teaching. Yet little is really known about what effects different kinds of questions have upon pupils' achievement or verbal responses (Gall, 1970). Very few studies have concentrated upon teacher questioning as such, and most of the recording of the kinds of questions asked by teachers has been done as part of a broader study, in which questions have been just one aspect.

In these circumstances, the question types identified have often been simplistic. Terms like 'broad' and 'narrow' have been common. There are only a few category systems designed exclusively to record the types of the questions asked by teachers, and even fewer systems have been constructed to record questions asked by teachers in social studies lessons.

In this section the purposes of teacher questions are examined and evidence is cited that teachers have tended to ask a restricted range of questions even though they have frequently been urged to do otherwise. A review is made of some of the types of questions that have been used in classroom research and particular attention is given to several major category systems and their theoretical derivations. This review is restricted to teachers' oral questions.

2.3.1 Purposes of teacher questions

Cunningham (in Weigand, 1971, p.83) defined a question as 'a verbal utterance that seeks a response from the person to whom it is directed'.

In pointing out that questions have always been the 'stock-in-trade' of teachers, Cunningham noted the purposes of teacher questions as giving directions, correcting and managing behaviour and activity, initiating instruction, creating learning situations, evaluating learning, and stimulating thinking. He viewed the use of questions,

by teachers, as a vital part of the teaching-learning process, since they give teachers the ability to develop desirable pupil attitudes and interests towards school studies, provide pupils with alternative ways of studying and thinking, and increase the quality of evaluating what the pupils learn. However, the purpose of teacher question-asking which is of prime importance, according to Cunningham, is to develop pupils' thinking.

Taba *et al.* (1971) agreed, but pointed out that the kinds of questions that need to be asked by a teacher depend upon the purposes of a discussion. For example, if pupils are required to recall facts from a previous discussion, then the questions will differ from those used in a situation where a teacher wants pupils to say what they think about the possible outcomes of a culture conflict situation which has been presented to the class.

This may partly explain why there is no clearly developed theory of teacher questioning. It was seen in the discussion on inquiry earlier that even though various proponents of inquiry might agree upon the philosophical issues involved in the method, when it came to specifics such as questioning, practises varied (for example, Suchman compared with Taba). The theoretical assumption that seems to be common — in relation to questioning — is that inquiry teaching should try to develop 'higher' levels of thought in pupils, and, that pupil achievement will thus improve.

At this stage of what is known about questioning, then, Taba's suggestion of analysing the kinds of questions according to the purposes of the discussion in which they were utilized might be the best approach. Gall (1970), too, regarded it as necessary first to set educational objectives for particular settings and then to consider what might be appropriate.

Regardless of the recognition that a major purpose of teacher questioning is to develop various thinking skills in children, however, there is ample evidence that this objective has often not been realised. It appears that most questions asked by teachers in class-rooms have been those that require pupils to use low-levels of thinking, and that children have been infrequently required to use higher-level thought in answering questions (Gallagher and Aschner, 1963; Gall, 1970; Ladd and Anderson, 1970; Wright and Nuthall, 1970;

Rosenshine, 1971; Wragg, 1974).

As long ago as 1912 Stevens (in Clegg, 1971) studied the role of questions in the classroom and found that in high-school English and social studies lessons the major emphasis was placed on the asking — by teachers — of memory questions. Two-thirds of the questions required direct recall of information. She urged that teachers should use a wider variety of questions to develop reflective thinking in their pupils. Two decades later Haynes found that 77% of teachers' questions in Grade 6 history classes called for factual answers and only 17% were judged to get children to think at a higher level (cited by Gall, 1970).

More recently, similar patterns have continued to be observed. At the elementary level Guszak (1967) found that recall questions made up 57% of the total questions asked by Grades 2, 4 and 6 teachers in reading and Schreiber (1967) found a similar pattern in social studies. After analysing teacher questions in several subject areas, Gallagher (1970) found that cognitive memory questions made up over 50% of the total questions in nearly every lesson. There were relatively few evaluative and divergent questions asked, and in some lessons these types were entirely absent.

It should be noted, however, that Gallagher and other researchers (for example, Tisher, 1970) have found marked variations in the number of questions of each type asked by different teachers even when using the same lesson materials. These differences have also occurred over a range of subject fields.

2.3.2 Types of questions

A large variety of types of questions have been identified by different writers and investigators. Most of these have been included within systems measuring categories of verbal behaviour.

It was noted by Dunkin and Biddle (1974) that, until recently, most measured variables of classroom behaviour were related to social-emotional characteristics. The most influential system in this respect has been the Flanders' Interaction Analysis Category system which was designed primarily to measure affective classroom behaviour (Flanders, 1970). But although this system has been widely used in studies, the category 'asks questions' has failed to

adequately differentiate and describe the complex field of questioning.

In an attempt to get more information about the cognitive, as well as affective, aspects of teacher questions the Verbal Interaction Category System of Amidon and Hunter (1966) included 'asks narrow questions' and 'asks broad questions' as two categories of teacher initiated talk from Flanders' system. The same authors (1967) then developed a modification to the Flanders system by including the Aschner-Gallagher categories (to be discussed below) under Category Four: 'Asks'. This gave the modified version scope to categorise teacher questions more specifically according to their cognitive elements: cognitive memory, convergent, divergent, and evaluative thinking. This resulted in a system with greater power in the analysis of teacher questions, even though it was still concerned mainly with affective variables.

These modifications reflect a growing interest by researchers over the last decade in looking more closely at the cognitive aspects of classroom behaviour, and have resulted in the development of a number of systems to measure cognitive variables, including teacher questions. Dunkin and Biddle (1974) saw these systems as having one or more of three theoretical orientations. Those with psychological orientation were concerned particularly with the nature of intelligence and with trying to identify intellectual skills and abilities being used in the classroom. A second orientation is seen in those studies that have measured the extent to which classroom behaviour is 'logical', and the third is embodied in studies involving verbal communication and its forms and sequences.

The review below will emphasise systems derived from the first orientation since the present study is primarily concerned with the cognitive aspects of inquiry.

Within the psychological, or cognitive orientation, Dunkin and Biddle (1974) identified three major cognitive systems that they believed have generated most devices for recording the cognitive behaviour in classrooms: Bloom's *Taxonomy of Educational Objectives* (Bloom, 1956), the cognitive operations developed by Taba and her associates (Taba *et al.*, 1964) and Guilford's Structure-of-Intellect model (Guilford, 1956).

Bloom's *Taxonomy of Educational Objectives* and the derived systems

were formulated on the assumption that classroom discourse could be sorted into categories that were related to a hierarchy of thinking levels: knowledge (the lowest level), comprehension, application, analysis, synthesis, and evaluation (the highest level).

One major system for classifying teacher questions, which was based on the Bloom taxonomy, is the *Teacher-Pupil Question Inventory* [TPQI] (Davis and Tinsley, 1968). The categories are as follows:

1. Memory — the person questioned recalls or recognises information (facts, generalizations, etc.);
2. Translation — the person questioned changes information into a different form (linguistic, symbolic, image, etc.);
3. Interpretation — the person questioned states relationships between various types of data;
4. Application — the person questioned solves a realistic problem requiring the identification of the crucial issue or points and the selection and use of appropriate knowledge and skills;
5. Analysis — the person questioned answers with explicit attention to the relationship(s) between the ideas expressed and with obvious awareness of the process employed in the reasoning;
6. Synthesis — the person questioned suggests answers to a problem that is original, speculative or creative;
7. Evaluation — the person questioned makes a judgment according to explicit criteria (external or internal);
8. Affectivity — the person questioned responds with a statement of feeling, emotion, or opinion without a standard of appraisal;
9. Procedure — the question relates to classroom organisation, student behaviour, or instructional management.

Another observational system which has used the Bloom categories is the *Florida Taxonomy of Cognitive Behaviour* [FTCB] (Brown *et al.*, 1968). This system lists 55 classes of behaviours which are distributed across the seven categories: knowledge, translation, interpretation, application, analysis, synthesis, and evaluation. The

FTCB is a sign system, however, and the authors divide observation time into six-minute periods, and only score a mark for a behavioural class the first time it occurs. Therefore, if one kind of question were asked several times in the six minutes, it would only be scored once, whereas the TPQI scores teacher questions every time they occur.

Gall (1970) and others have cautioned that there are limitations in trying to use the categories of the Bloom (1956) *Taxonomy of Educational Objectives* to classify teacher questions. Some categories are hard to distinguish from others, as they depend upon inferences regarding the cognitive behaviour involved. Thus when a question is answered by a pupil, it is not always easy to determine what precise level of thought was required to make the response. On the other hand, Sanders (1966) claimed that Bloom's categories would satisfactorily differentiate between different kinds of teacher questions and that the categories were applicable to the analysing of questions at all age levels of the school, and for all subject areas. But Sanders did not actually try out the question types by analysing classroom discourse!

In addition, because the Bloom categories are often difficult to distinguish, it would be difficult to teach the questions related to them to teachers. Therefore the use of such categories in short in-service programmes is probably limited.

A further disadvantage of the Bloom categories is that the questions developed in the above systems are of the initial kind that a teacher uses as a lead-in. No specific provision is made for teacher questions which take initial answers further. Also, no provision is made to record classroom functioning. For example, in inquiry lessons one of the challenges for teachers is to maximise pupil response: a detailed recording of managerial functions may give some indication of how far this is being achieved.

The cognitive operations developed by Taba and her associates (Taba *et al.*, 1964; Taba, 1966) have been a second major influence. The operations identified by Taba and incorporated in her category system are: concept formation, inferring and generalizing, and application of principles. The coding system used in these studies was designed to record the thought units and thought levels of each utterance in class discussion. Teacher questioning was considered

very important, because Taba believed that pupils usually follow the teacher's lead, so that if a teacher asks for a fact, then the pupils usually give him a fact. Therefore, the teacher 'tends to set a ceiling on the level of responses that are given' (Taba *et al.*, 1964).

Taba's interaction system, consequently, included categories designed to identify attempts to 'lift the level of thought' in pupils through question-asking, such as 'teacher seeks inference, generalization, or principle', and 'teacher seeks explanation'. The question types cover a wide range of thinking abilities and, as such, seem appropriate for research into inquiry teaching which tries to get children to think at a high level. However, the full system is perhaps rather too complex to be taught to teachers in relatively short in-service programmes without follow-up training activities. Nevertheless, the Taba system has been extensively applied to social studies lessons, and will be referred to in more detail later.

A few studies have been based on the third major system: the Guilford Structure-of-Intellect model (Guilford, 1956). The orientation of this group has been directed towards 'divergent' thinking in classrooms. The Aschner-Gallagher system (Gallagher, Aschner and Jenné, 1967) used the primary thinking categories: cognitive memory, convergent, evaluative, and divergent. A fifth category 'routine', covers managerial functions. These five categories were subdivided into complex sub-categories, into which teacher questions are analysed.

The full system is more appropriate for research purposes than applied to in-service work with teachers, because there are over fifty sub-categories. This not only makes coding a problem, but many hours of training would be necessary to familiarise teachers with the categories, and to use them in analysing their own lessons.

If, however, the system is simplified it seems ideally suited for in-service programmes where teachers need to be taught the questions in a relatively short time. Also, the Guilford thinking operations can be related to a range of instructional objectives appropriate for inquiry lessons.

With in-service education in mind, Cunningham (1971) developed a modified version of the Aschner-Gallagher system. He used the five basic categories but added Amidon and Hunter's 'narrow'and 'broad'

questions. Cognitive memory and convergent questions were considered narrow and evaluative and divergent questions broad.

While Cunningham's system appears to be well suited for use in in-service programmes, like most other systems cited so far it includes only initial-type questions and makes no reference to what a teacher does with answers he gets from his pupils.

It should also be noted that Gallagher (1965) found an inter-relationship between the thinking operations which sometimes made distinctions between them difficult when coding lesson transcripts. This was a particular problem with convergent and divergent thinking.

It was pointed out earlier in this chapter that theorists have given a good deal of attention to the use of logic in instruction. Only recently though has any systematic study been carried out. Smith and Meux (1962) developed a complex system for analysing the logical structure of teaching. Classroom discourse was measured by 'episodes' which comprise a statement (which might be in the form of a teacher question), a continuing phase, and a terminal phase. The cognitive operations involved in the demands made on pupils were categorised as follows: defining, describing, designating, stating, reporting, substituting, evaluating, opinion, classifying, comparing and contrasting, conditional inferring, and explaining. Procedural comments were included also.

Nuthall and Lawrence (1965) modified Smith and Meux's episodic analysis by using incidents as the basis for determining the logical structure of verbal moves. The incident consists of any question or demand and all the subsequent verbal moves that result. Of particular interest is that Nuthall and Lawrence distinguished between 'major' questions and 'subsidiary' questions. Subsidiary questions were follow-up questions to a prior (major) one. It was noted above that a disadvantage of other systems is that no provision is made for differentiating follow-up questions.

It would seem that incident analysis could be a useful in-service device for getting teachers to study their own questioning behaviour.

The questions categorised by Nuthall and Lawrence (1965, pp.27-8) from language lessons were:

1. Vague questions: difficult to tell what is required, e.g., 'Once a girl hears "sewing machine" what does she want to say next?'

2. Questions requiring opinion and expressing of feelings: e.g., 'Who'd prefer the Necchi?' 'Don't you agree?'

3. Closed questions:
   - 3.1 Requiring a specific factual answer directly related to the pupil's experience: e.g., 'What sort (of sewing machine) has Mum got?'
   - 3.2 Requiring simple 'Yes' or 'No' answer: e.g., 'Is it (made of) metal?'

4. Open questions:
   - 4.1 Requiring some reference, implicit or explicit, to the characteristics of class, or to the explanatory principles: e.g., 'What is it that makes a machine?' 'Why is it (a radio) a machine?'
   - 4.2 Requiring an example of or instance of a class or type of object or answer: e.g., 'Can you tell me of a machine that does not have one of these requirements?'

5. Repeated questions: repeating substance of previous questions in similar wording.

The Ford Teaching Project developed a system so that teachers could self-monitor their classroom questions (Elliott and Hurlin, 1975) based upon a number of hypotheses and ground rules, relating to inquiry/discovery teaching in a variety of subject areas. These authors recorded a number of lessons in history, and identified question types used by teachers, which are summarised as follows:

1. <u>Recall questions</u>

   a. Questions that assume the correct answer to be forthcoming.

   b. Questions seeking information which leaves the resources open to question (e.g., 'Where did you find that?').

   c. Questions seeking information about how the pupil proceeded (e.g., 'Are you guessing or did you read that?').

   d. Subject-centred questions, focussing upon the lesson content.

   e. Real questions, which seek information (this is used informally).

   f. Pseudo-real questions, which involve a pretence of ignorance by the teacher in order to find what the pupil knows.

   g. Exam questions to test if the child really knows something.

   h. Pseudo-exam questions, which involve pretence by the teacher to test knowledge.

2. Deductive questions

   a. Questions requiring reasoning about information in the lesson resources.
   b. Modified assertions.
   c. Question sequences which get children to reason out conclusions.

3. Evaluative questions

   a. Inquiry sequences to help pupils clarify and explore their own ideas.
   b. Pupil-centred questions; pupils' ideas about subject matter.
   c. Personal opinion questions.
   d. Questions that ask pupils to put themselves in another person's position.
   e. Moral judgment questions about right and wrong.
   f. Modified moral judgment questions which extend the previous category.
   g. Questions which focus on conflict of ideas in the resources.

The above categories were not intended, by their formulators, to be definitive. Instead, their intention was that teachers were to be encouraged to try out the various question types to see whether they (the teachers) felt that any improvement in teaching resulted. Therefore the question classifications were regarded by the researchers as tentative, and they intended to get teachers to suggest modifications. It should be noted that some of the categories are sequences rather than questions in their own right. To date, no results of the usefulness of these question types have been reported but they do appear to include — collectively — a wide range of thinking possibilities for pupils.

2.3.3 Questioning in social studies classrooms

In social studies, as has been the case with other subjects, the consistent finding of research studies is that questions by teachers in social studies lessons mostly require only low-level thinking by pupils. In general, they call for little more than the exercise of memory recall (Aschner, 1961; Gallagher, 1965; Hunkins, 1966; Taba, 1966; Davis and Tinsley, 1968; Gall, 1970; Godbold, 1970; Tinsley and Davis, 1971; Rogers, 1972; Chaudhari, 1975).

For example, Godbold (1970) used Bloom's taxonomy to classify the questions asked by two groups of primary school social studies teachers and two groups of secondary teachers of social studies. The percentage of memory questions ranged from fifty-five to sixty-eight per cent of all questions asked. No analysis questions were posed by any teacher, and primary teachers asked, on average, ten per cent more memory questions than did secondary teachers. Length of teaching experience did not influence the frequency of use of question types.

Written lesson questions asked by secondary school student teachers of social studies were studied by Tinsley and Davis (1971). The majority of these questions were either memory or evaluation (as measured by Aschner-Gallagher categories). Convergent questions were also stressed, though to a lesser extent, but few divergent questions were asked.

Although this review is concerned with oral questions, it is of relevance that in units of work in textbooks specific questions are often suggested for teacher and pupil use, and these also give an indication of the level or oral questioning adopted by teachers. Davis and Hunkins (1966) found that seventy-eight per cent of the questions in social studies textbooks they reviewed were at the memorisation level. Of the 732 questions they examined, none required thinking at the synthesis level, and only two required evaluative thinking. Chaudhari (1975) also found that the predominant type of teacher question printed in textbooks in all subjects simply required pupils to recall information; he found little evidence that textbook questions explored the higher levels in children's thinking.

Consequently, it is not surprising that researchers have recently been exploring different kinds of questions that might be appropriate for fostering and developing inquiry learning in social studies, with the intention of widening the repertoire of question types used by teachers. However, to date, very few systems have been developed which classify questions specifically with social studies lessons in mind.

One such system, however, was designed for social science teaching as part of doctoral research by Schreiber (1967). It included the following categories:

Recall: recall of facts; arranging facts in sequential order.

Analytic thinking: making comparisons; identifying supporting facts; drawing conclusions.

Creative thinking: speculating on outcomes.

Evaluative thinking: identifying main parts and important parts; stating moral judgments; stating judgments based on personal experience; evaluating quality of source material; evaluating adequacy of data.

Other: describing situations; defining and clarifying information; using globes; using maps; uncovering information and raising questions for study.

Schreiber's system, which she intended to be appropriate for inquiry teaching, has clear relationships with the major category system of Aschner and Gallagher (Aschner *et al.*, 1965), in that it is based on similar cognitive elements. However, the only provision made for additional categories in which a pupil response to a question is followed by a further teacher question to achieve thinking beyond the initial response, is the 'supporting facts' category.

Another system for classifying teacher questions in social studies was developed by Ryan (1971). The categories were: recall, relationship, application, educated guess, synthesis, and opinion. Five of these levels were designed to record questions requiring pupil thought above the memory recall level. The 'educated guess' category was intended to be 'specially amenable to discovery, involvement, and inquiry strategies ...' (*ibid.*, p.123). This category calls for inferential thinking; for example: 'What would happen if vast supplies of oil were discovered one hundred miles north of Las Vegas, Nevada?' (*ibid.*, p.124). This is a similar type of question to those used in the 'Just Suppose' items in the *Torrance Tests of Creative Thinking* (Torrance, 1966). Ryan's system was thus designed to take note of those higher-level questions appropriate for inquiry teaching, but he stressed that such a question system would only be useful to teachers in an appropriate, open-ended classroom climate. Taba (1966) and Flanders (1967) also emphasised that for a variety of question types to flourish, classroom climate must be conducive; and that the climate is set by the teacher. Thus Taba *et al.* (1964) warned that, even if numerous open-ended questions were asked in social studies lessons, children might still be fearful of saying anything because they were not sure if it was what the teacher wanted to hear them say. Somehow,

therefore, a teacher has to communicate the message that pupil opinions and original thinking are really what he wants.

The asking of genuine open-ended questions was thus viewed by Taba as particularly important in elementary social studies teaching. It was pointed out that, in class discussion of topics, a series of closed questions inevitably produces an interaction sequence of teacher-pupil-teacher, whereas open questions can encourage numerous children to participate in response to perhaps just one question (Taba *et al.*, 1964).

As mentioned earlier two types of questions identified by Taba and her associates were 'lifting questions' and 'supporting questions'. The former is a type of extension question, and was seen by Taba to be important for social studies teaching because it gets children to think beyond initial answers with questions like 'What might be the result of that?' The type of extension question — 'supporting' — was considered significant because it allows the teacher to 'read' feedback from pupils as they answer questions, and then make additional requests to develop thinking further, either in the pupil who gave the initial answer, or in another student. McNaughton *et al.* (1967) also found that extension and justification-type questions helped children reach sound conclusions.

The research of the Taba Project (Taba *et al.*, 1964) represents the major attempt to date to relate teacher questioning to cognitive theory in social studies. Taba believed that pupils' thinking could be developed through systematic questioning which successfully lifted and extended through to higher levels of complexity. An extension of this work was carried out by Wallen *et al.* (1969) in which the researchers measured effects of teachers' attempts to develop student thinking skills upon pupil learning (on several achievement tests). The results are discussed further in Chapter 3, but it can be noted here that they demonstrate that the Taba system can be taught to teachers, and that the categories cover a wide range of thinking abilities in pupils.

Four question categories were suggested by Massialas (in Ploghoft and Shuster, 1971, p.223) as being useful for secondary school teachers specifically in discussions of social issues. They were:

1. Questions that call for expository responses.
2. Questions that require definition and clarification.
3. Questions that get pupils to give positions or offer hypotheses.
4. Questions that call for grounding.

Ploghoft and Shuster themselves regarded these categories as suitable for elementary school, as well as secondary school, social studies lessons. The 'system' does, however, seem still to stress convergent thinking, rather than to encourage the development of pupils' opinion-giving skills, or extension of thinking.

Apart from Taba's category system of questions, little is known about whether the systems above can be readily taught to teachers. It would appear, then, that there is a need to explore the suitability of various categories of question for social studies inquiry teaching, and to see whether they can be learnt easily by teachers. The urgency for this has been clearly shown by the finding of Bellack *et al.* (1966) that in sixteen secondary social studies classrooms, there was one pattern of discourse that dominated: 'teacher solicits, student responds, teacher reacts'. A pattern like this places severe limitations upon the possibility for children to engage in complex thinking and open-ended debate.

2.3.4 Problems of research on teacher questioning

One of the greatest problems facing researchers in the field of recording, analysing and comparing the results of teachers using different types of questioning is that there is no simple agreed system of categories or even of labels for essentially similar categories in different systems. For example, the terms 'open' and 'closed' have been used by Hough and Duncan (1970) to distinguish between convergent and divergent thinking; by Wright and Nuthall (1970) to separate relatively direct questions which could be answered without the need for additional information from questions which involved more complex thought; and by Cameron-Jones and Morrison (1973) to distinguish between those questions which do not contain teacher-given clues and those in which he does give clues to the pupils as to the response he expects.

The resulting difficulty in comparing experimental findings has

been pointed up by Heath and Nielson (1974); they have claimed that Rosenshine and Furst in their earlier review (1971) of process-product research, had incorrectly grouped variables as similar in a conceptual sense when there was in fact little similarity. For example, there is indeed little possibility of valid comparison between the findings of the studies by Hough and Duncan, Wright and Nuthall, and Cameron-Jones and Morrison, as to the effect of open and closed questions on pupil behaviour, because they are talking about differently-classified instances of verbal discourse.

There would seem to be two ways of overcoming problems of comparability. One way would be for investigators to relate the variables to a common 'conceptual field', such as a particular theory about the cognitive functioning of children (for example, Taba, 1966). Another would be for all researchers to formulate clearer operational definitions for the question types they are recording.

The degree of inference is an important consideration when teacher behaviour is analysed. For example, measures are of low-inference, according to Rosenshine (1971, p.19), if the

> items focus upon specific, denotable, relatively objective behaviours such as "teacher repeats student ideas" or "teacher asks evaluative question", and also because these events are recorded as frequency counts.

On the other hand, high-inference measures lack the specificity of low-inference variables. Rosenshine (*ibid.*) explained this in the following way:

> Items on rating instruments such as "clarity of presentation", "enthusiasm", or "helpful towards students" require that an observer infer these constructs from a series of events. In addition, an observer must infer the frequency of such behaviour in order to record whether it occurred "consistently", or "never", or somewhere on the set of gradations used in the scale of the observational rating instrument.
>
> ... the low-inference statement that a teacher repeated student answers seven per cent of the time is much more specific than the statement that a teacher was "sometimes helpful towards students".

The categorising of skills such as teacher questioning would therefore be low-inference measures, provided that the question types were clearly defined for coders. Greater precision is also achievable if

skills can be coded from transcripts instead of the observer having to make inferences as a lesson proceeds, as is the case with some systems. More thought can, as a result, be given by the coder to what preceded the teacher question, and what followed, so that the type of question can then be more reliably determined.

In some studies, teacher questions have been classified into more than two categories, and the categories combined into a ratio for statistical analysis, which meant that the analysis related, in effect, to a single variable. For example, Allen (in Rosenshine, 1971, p.126) used a modification of the Bloom (1956) taxonomy to classify questions as knowledge, translation, and interpretation and application. In one ratio, knowledge questions formed the denominator, and the other question types were in the numerator. In a second ratio, interpretation and application were in the numerator. Neither ratio revealed significant relationships with pupil achievement in eighteen Grade 1 mathematics classes.

One problem in this kind of analysis is that the effects of specific behaviours can be obscured by statistical analyses in which the specific categories are combined. Rosenshine (1971) indicated that the use of ratios has interesting possibilities in providing information about classroom behaviour effects, but he recommended that analyses of the individual behaviours (for example, different kinds of teacher questions) should always be carried out in addition to the use of ratios.

2.3.5 Summary

Teacher questioning skills have been clearly recognised as a very important part of inquiry learning and teaching by numerous writers (for example, Bruner, 1966; Taba, 1966; Massialas and Zevin, 1967). While many kinds of questions have been suggested, there is no general system for classifying questions for all inquiry teaching. Indeed, it may not even be appropriate that there should be, for it has been suggested (Taba, 1966; Gall, 1970) that the formulation and use of question types should be carried out in relation to particular objectives for particular educational settings. That this has, in fact, occurred on occasions was shown in the previous section when various inquiry strategies (Suchman, 1961; Bruner, 1966; Massialas and Zevin, 1967) were presented. At the same time, however, it seems

clear that there are certain kinds of questions which are probably appropriate for teaching generally.

Researchers over a long period have demonstrated that teachers in various settings have employed a preponderance of questions which require little thought from their pupils, in spite of the clear recognition that the development of various thinking abilities is a major educational goal. There would appear to be a need for researchers to formulate systems of questions for use in particular educational settings and which include categories covering a wide range of cognitive objectives.

It would also seem to be necessary, in looking at true discussion-type lessons, to consider two kinds of questions:

1. Initial questions which engage children in a wide range of thinking abilities. The operations categories of Guilford's intellect model would appear to be suitable bases for such questions, as has been demonstrated by Gallagher (1965).

2. Follow-up questions which encourage children to develop responses to initial questions. Some of the terms identified in past studies which describe these questions are 'probing', 'extension', 'grounding', 'lifting', and 'elaboration'.

Many of these question types which have been identified in this section have been developed within systems designed to record more than just teacher questions. Furthermore, most of the systems themselves have been used for research purposes; that is, for recording the sorts of verbal behaviour teachers do use, rather than for training teachers how to improve their classroom communication skills.

In-service education is now increasingly recognised as a way of achieving such improvement (Gall, 1970; Cunningham, 1971; Borg, 1975). Not only can teachers be encouraged to build-in new questioning approaches; they can also be trained to eliminate 'problematic' questions. For example, Gall (1970, p.710) urged researchers to give attention to teaching these skills:

1. Questions which cue students to improve on an initially weak response to a question ('Can you tell me a little more?'; 'What do you mean by that?').

2. Questions which create a discussion atmosphere ('Billy, do you agree with Sue's position?').

3. Questions which stimulate students' sense of curiosity and inquiry ('What would you like to know about this manuscript?'; 'How would you propose to find an answer to this question?').

4. Questions which guide students' learning of a problem-solving or affective skill ('What do you think we do next to solve this problem?'; 'Mark, what is your response to these drawings?').

Also, Cunningham has suggested teaching teachers to ask fewer 'yes' and 'no' questions because asking too many in a lesson may inhibit pupils' thought. Wright and Nuthall (1970) have suggested that ambiguous questions have the same effect.

In the following chapter some in-service programmes which have been tried to change teacher questioning behaviour will be described, and an analysis will be made of the effects of teacher questioning upon pupil achievement, and the relationship between changes to teachers' questioning behaviour (resulting from in-service programmes) and pupil attainment.

CHAPTER 3

IN-SERVICE COURSES AND THEIR EFFECTIVENESS

OVERVIEW: Doubts about the effectiveness of generalised in-service courses have led to a shift in emphasis towards more systematic approaches, in which specific teaching skills have been identified and subsequently taught to teachers. The goals and skills of teacher questioning that have been formulated in different in-service projects are outlined, and evidence cited that such programmes can change teacher behaviour. Consideration is then given to how pupil attainment has been influenced by teachers' question-asking behaviour, and especially to 'second-step' effects resulting from in-service training.

3.1 In-service and pre-service training: some new goals

As pointed out in the introduction to this thesis, there has been a growing awareness that teacher education, both pre-service and in-service, is not as effective as it might be. In a survey of research into the process of teacher education Peck and Tucker (1973) conclude that only in the last few years has much worthwhile empirical research been carried out to investigate teacher education programmes that try to train teachers to develop specific skills. Part of the reason is that skill-teaching is often a long, complex step-by-step process, and many of the relevant skills, such as questioning, have not been fully analysed nor performance accurately measured. Perlberg (1972, p.553) concluded: 'In spite of their potential importance, in-service education programmes have little effect on teacher-pupil interaction'.

Similarly, after reviewing literature related to in-service teacher education, Borg (1975) concluded that these courses had largely failed to produce changes in teacher performance. One reason suggested by Borg for this failure was 'that these courses tended to deal with generalities rather than identifying specific behaviours that teachers could employ to bring about specific outcomes' (*ibid.*, p.305). Another failing he noted was that 'most courses were taught primarily using

lecture and discussion techniques' (*ibid.*). Consequently, teachers lacked effective models and feedback, and there was a lack of focus on specific skills.

To counter such criticisms several new approaches have been suggested over recent years for increasing the effectiveness of in-service teacher education. These include training teachers in the general techniques of analysing teaching situations; utilizing a 'systems approach' incorporating feedback procedures; using techniques such as microteaching to develop specific skills and competencies; and training teachers in the use of specific skills required to reach particular curriculum objectives.

One of the most influential of these new approaches has been that of Flanders (1967; 1970), who developed what was initially a research technique into a training procedure to help teachers decrease their reliance on direct control of classroom activities and increase their acceptance of increased pupil participation. His method was to train teachers in the use of his FIAC system so that they could both become more sensitive to what they were actually doing in the classroom and also get relatively immediate feedback on their attempts to change their teaching behaviour. Others, such as Amidon, Hough (Amidon and Hough, 1968) and Hunter (1972) developed pre-service and in-service programmes from this base.

Other programmes have incorporated the systematic analysis of researchers such as Smith *et al.* (1964) to direct attention to the cognitive implications of particular classroom verbal forms and behaviours. Thus in planning the Utah State University Protocol Project, Borg (1975) sought to develop skills and materials that would train teachers in specific behaviours for use in their own classrooms. His materials were developed from ideas provided by Smith *et al.* (1969), who considered protocols to be most effective if they were based on original recordings of classroom interactions which could then be used to help teachers relate important educational concepts (for example, planning) to actual classroom events. Borg's model contained the following steps for the learner:

1. Scan the learning sequence to get a step-by-step outline of what he will do.

2. Read the module objectives, a description of the concept, and the specific teacher behaviours to be used to apply the concept in the classroom.
3. Complete Recognition Practice Lessons from transcripts of classroom audiotapes, to identify instances of the behaviours being learnt.
4. View the Protocol Film and identify instances of the behaviours being learnt.
5. Do a Recognition Test, designed to measure the learner's ability to recognise classroom applications of the behaviours being learnt.
6. Complete Application Practice Lessons from classroom audiotapes. The learner would write in appropriate remarks that apply to the teaching behaviour at points in the class discussion where the teacher's remarks have been deleted from the transcript.
7. Plan a brief lesson designed to practice the teaching behaviours. Teach the lesson, and taperecord it.
8. Replay the lesson with another teacher and record the use of the specified behaviours on a tally sheet, and discuss.

A 'systems' approach was recommended for teacher education by Peck and Tucker (in Travers, 1973, p.943). Their generalized programme consists of a series of steps which recur in cyclical fashion as follows:

1. Precise specification of the behaviour which is the objective of the learning experience;
2. Carefully planned training procedures aimed explicitly at those objectives;
3. Measurement of the results of the learning in terms of the behavioural objectives;
4. Feedback to the learner and the instructor of the observed results;
5. Re-entry into the training procedure (a trial-teaching experience, for example);
6. Measurement, again, of the results following the repeated training.

This general model has been used for training teachers in interaction analysis, microteaching, and behaviour modification.

A model of in-service teacher education based on four stages was developed by Mohan and Hull (1975): a planning stage; an implementation stage; an evaluation stage; and a follow-up stage. In the evaluation stage, the authors recommended the use of cognitive tests and questionnaires to measure feelings and reactions, and to get participants to identify programme limitations. A checklist of teacher performances was recommended as a follow-up to the programme. Teachers could then use the checklist to evaluate their own performance of the competencies dealt with in the in-service course.

A seven-point in-service programme was put forward by Olson (1974). The components are: a well-defined set of goals and materials; diagnostic materials; alternative learning strategies; a learning environment for flexible usage of the strategies; the set of teacher competencies; evaluation through sharing of ideas; and a monitoring system to check progress in competency learning.

This model was designed for use in individualised in-service education, and when a teacher successfully completes one competency, this is noted in the computer-assisted monitoring system, and the teacher proceeds to the next competency. What this model does not allow for, however, is any personal contact with colleagues, to discuss and evaluate skills development. Also, it is doubtful whether the specific skills concerned can be learnt in isolation from other skills.

A whole generation of pre-service and in-service training programmes has resulted from the development of microteaching procedures, initially at Stanford University but later in one form or another at many institutions in other parts of the world (see, for example, Turney *et al.*, 1973). The common feature of all such programmes is an emphasis on the identification and practise of specific teaching skills, either in real or simulated situations. Programmes based strictly on the Stanford model are behaviouristically oriented, while others have adopted a more holistic approach (Freyberg, Katterns and Rogers, 1974). Thus Brown (1975), for instance, has formulated a microteaching programme which focusses upon specific skills, but tries to relate these to a more general model of teacher education. This stress is important, because teachers can be encouraged to learn specific behaviours, yet are enabled to see the

wider context in which they are used. The broader type of programme formulated by Brown included being aware of the planning of lessons (selecting objectives and methods of teaching them), performance (acting-out the skills), and perception (alerting the teacher to his performance and making modifications to further subsequent planning).

Probably the most extensive training programmes yet designed to teach teachers how to improve the thinking abilities of children were constructed under the leadership of Taba (Taba, 1966). These were based on three propositions about the learning of new or different skills. First the learner should understand the rationale and structure of the curriculum. Second, the learner should be given examples of the nature and function of the learning experiences. Third, the learner must be provided with full details of the teaching strategies necessary for the acquisition of the skills involved.

The Taba project emphasised that it is not sufficient to produce a number of teaching skills as such. Rather, the skills should be set in the context of the curriculum being taught.

A training period of ten days was used in the Taba project. Taba suggested that this period was sufficient for immediate teacher change, provided three criteria were met. First, the training should provide a sound theoretical framework for the strategies in terms understood by teachers. Second, examples of the strategies must be presented. Finally, immediate feedback about their performance should be provided for teachers (Taba, 1966).

The manner in which such skills are taught to teachers is also of crucial importance. Rubin (1973) suggested that in-service educators must distinguish between inner self-determinism and outer manipulation. On the one hand a teacher may decide to alter his behaviour, and on the other hand the teacher may be coerced. The former is likely to be more effective in the long-term, because self-decision is more likely to produce a flexible person who will change when circumstances are appropriate for change.

From this brief outline of the approaches and models mentioned above it can be seen that their authors emphasise several important elements in designing in-service courses with the intention of changing specific aspects of teachers' behaviour. These are:

(a) The teaching skills concerned need to be clearly identified.

(b) Teachers need to understand how these skills can be used.

(c) Opportunity to practice the skills must be given.

(d) Feedback about performance should be provided.

(e) Modifications that are necessary as a result of the feedback have to be made by the individual teacher, so that the emphasis should be on teacher decision-making as well as on specific skills.

3.2 Changes in teacher questioning behaviour

A number of studies have demonstrated that teachers' use of question-asking skills can be changed through systematic in-service training programmes (for example, Ryan, 1969; Zevin, 1969; Borg *et al.*, 1970; Merwin and Schneider, 1973; Galassi *et al.*, 1974).

The main goal of such experiments has been to encourage teachers to reduce the number of questions which call for low-level thinking responses in pupils, and to increase questions which involve higher and more complex levels of thought. In an early study directed towards this goal, Houston in 1938 (in Gall, 1970) developed an in-service programme which made use of group conferences, self-analysis, and supervisory conferences to promote change in teachers' questioning procedures. He reported that the programme resulted in teachers using more higher cognitive questions, more questions relevant to the subject matter, and more student participation.

The continuing need for this increased use of higher-level questions was emphasised by Gall (1970) in a review article. Gall reported that in studies he reviewed the mean percentage of higher cognitive questions asked by teachers without specific training was less than twenty per cent of all questions asked. Galassi *et al.* (1974) pointed out that this situation is partly the result of teachers' lack of knowledge about how questions promote learning and thinking, and that there is no known optimum balance between fact-seeking and higher cognitive questions to maximise learning. Nevertheless, 'it seems likely that teachers should make substantially greater use of higher cognitive questions than they presently do'

(Galassi *et al.*, 1974, p.16). Chaudhari (1975) blamed textbooks for perpetuating the preponderance of low-level questions usually found in most lessons.

It was pointed out above that in recent years, microteaching has become an accepted technique of training teachers in the use of specific teaching skills, such as teacher question-asking. Frequent use has been made of videotape as a feedback device to show teachers models of certain behaviours. Galassi *et al.* (1974) attempted to clarify the effects of perceptual (videotape) and written modelling in changing teacher behaviours, including questioning. The treatment involved about seven hours of intensive in-service training. Teachers were randomly assigned to either a video version, or a written version of *Minicourse Nine: Higher Cognitive Questions* (Borg *et al.*, 1970). The treatments were the same in content, but the difference was that whereas one group viewed videotaped models of skills, the other group read transcripts of the same skills. The course spanned four weekly lessons. Pretreatment and posttreatment twenty-minute lessons — in social studies — were analysed by raters. Questions asked by teachers were analysed according to Bloom's (1956) cognitive levels. Pupil behaviours were also measured by the number of student responses of a high-level nature, and length of pupil response.

Galassi *et al.* found that both treatment groups asked significantly more higher cognitive questions than a control group. But there was no significant difference between the two treatment groups. In asking probing questions, there was a significant increase for 'written group' teachers, and no significant difference between 'video group' and control group teachers. At both elementary and junior high school levels, 'written group' teachers were superior to 'video group' teachers in the number and in the length of higher level responses from pupils.

The researchers showed surprise at the superiority of the teachers who studied the written models of the teaching skills. They raised the possibility that the video group might have been distracted from the central purpose — practising teacher questions — by the unusual use of television. They thought that there was the possibility that the group studying the written version was better able to pace their learning. Also, it was pointed out that it was perhaps the process of developing a good model lesson for the videotape and transcription that

was effective (Galassi *et al.*, 1974, p.22). Incidentally, another interesting finding resulting from this study was that the pre-treatment percentages of higher-cognitive questions were higher than in most other reported studies. Galassi *et al.* suggested that supplying observation lesson materials to the teachers helped, especially as the materials were considered conducive to the asking of higher-cognitive questions.

In *Minicourse One* the Far West Laboratory programme focussed on twelve questioning behaviours of elementary teachers. These are:

A. Increase pupils' readiness to respond to discussion questions by:
   1. Asking questions; pausing three to five seconds; calling on a pupil.
   2. Dealing with correct answers in an accepting, non-punitive manner.
   3. Calling on both volunteers and non-volunteers in order to keep all pupils alert, and distribute participation.

B. Increase the amount of pupil participation, and decrease the amount of teacher participation by:
   4. Asking longer questions by:
      a. Asking for sets or groups of related facts when formulating information-level questions.
      b. Avoiding 'yes' or 'no' answers.
   5. Redirecting the same question to several pupils.
   6. Framing questions that require pupils to use higher cognitive processes.

C. Increase teacher use of <u>probing</u> techniques in order to guide the pupils to use more complete and thoughtful responses by:
   7. Prompting.
   8. Seeking further clarification and pupil insight.
   9. Refocussing the pupils' response.

D. Reduce teacher behaviour that interferes with the flow of the discussion by observing the following rules:
   10. Teacher should not repeat his own questions.
   11. Teacher should not answer his own questions.
   12. Teacher should not repeat pupil answers.

(Borg *et al.*, 1970)

As part of the training the teachers identified each skill behaviour, then made videotape recordings of their own microteaching practice, and through self-feedback and video models, measured their own teaching with the aid of checklists. Twenty-four of the original forty-eight teachers were followed through in a longitudinal study. They were recorded four times: pretreatment; immediate posttreatment; after a four-month delay; and again after a thirty-nine month delay. Three behaviours considered undesirable were permanently reduced to a significant extent. These were 'repeating the question', 'repeating the pupil's answer', and 'answering one's own question'. Teachers' use of higher cognitive questions in these Grade Four, Five and Six classes increased from 38% (precourse) to 50% (postcourse), and this level dropped only slightly over three years. It was found that the average proportion of time that teachers talked was 53% in precourse lessons, and 33% in postcourse lessons. However, after three years the teacher talk had increased again to 45%, therefore indicating a strong tendency for teachers to want to dominate classroom talk (Borg *et al.*, 1970). This programme is interesting because it not only encourages teachers to adopt new skills, but to eliminate undesirable ones.

The Waikato Teacher Education Research Project took account of the principle of 'self-decision' mentioned earlier when it developed an in-service training programme incorporating a number of cognitive teacher behaviours assumed to be important in the teaching-learning process in classrooms (Freyberg, Katterns, and Rogers, 1974). Three of the skills of significance for the present investigation were:

1. Structuring of memory recall questions for review purposes (equivalent to the knowledge category of the Bloom Taxonomy (Bloom, 1956) and Guilford's cognition and memory operations (Guilford, 1959)).
2. Asking questions to elicit high-order thinking responses (these were equivalent to the Bloom categories of comprehension, application, analysis, synthesis, and evaluation, and the Guilford operations: convergent, divergent and evaluative).
3. The use of <u>probing</u> questions to extend individual responses, which includes prompting, clarifying,

justifying, redirecting, and refocussing.

To teach the above (and other) skills, a microteaching programme was developed, which involved teachers in simulation exercises in ten half-days of in-service experience. Baseline data were gathered from pretreatment lessons, and posttreatment lessons were taught by the sixteen teachers two months after the in-service work. The post-observation lesson showed that the mean frequency of both memory recall and high-order questions decreased. However, the researchers also found that, although fewer high-order questions were asked, techniques such as probing and waiting for children to think, seemed to result in the pupils producing far more complex and thus lengthy answers. It might be assumed, therefore, that the quality of answers had increased. Certainly, the proportion of pupil talk was markedly higher in the follow-up lessons. Frequency counts alone, then, may not be a sufficient method of analysing changes in teacher skills. Also, 'wait time' would seem to be an important — though difficult — variable to analyse.

As in the research of Borg *et al.* (1970) reported above, self-instructional modules have been designed by Merwin and Schneider (1973), and the effects of these upon teacher trainees in social studies have been measured. Four modules each required three to five hours instruction time. The modules were expository, and included skills such as the use of higher cognitive questioning. The sixteen trainees of the experimental group were measured against a control group which was given a conventional course. Results showed that the module student teachers asked a greater number of higher-level cognitive questions subsequent to training. The authors concede, however, that the Hawthorne effect may have influenced the results (Merwin and Schneider, 1973, p.17). Even so, the study shows that changes can be produced in fairly short time.

Little is known about whether the effects of in-service training in questioning skills in one curriculum area will transfer to another. To test this Porterfield (1974) investigated differences in questions asked in reading lessons by teachers who were given in-service training in a particular science method of inquiry/discovery. The Teacher Question Inventory of nine question types designed by Harris and McIntyre was used to classify the questions of sixteen Grade 2

and 4 teachers and sixteen control group teachers (who had no science method training). The experimental group teachers asked a significantly larger proportion of questions at a higher cognitive level, and also asked more questions directed to pupils' beliefs and opinions. The questions asked by control group teachers were predominantly at the recall level. Porterfield (p.592) speculated that pupils' understanding was being limited by this tendency. This research indicates that transfer does take place from an instructional course which employs both theoretical ideas and discussions, into classroom behaviour in subject areas other than social studies.

Two studies which also showed that what is covered in an in-service course can transfer into teachers' classroom behaviour in social studies were those by Zevin (1969) and Ryan (1969). Both of these researchers ran in-service programmes in which the instructors tried to teach in a way that reflected what was termed 'inquiry' teaching. By this was meant a tendency to exhibit 'indirect' behaviours more than 'direct' behaviours as measured by the categories of Flanders' (1967) interaction system. Zevin and Ryan worked with experienced secondary teachers and inexperienced elementary teachers respectively, and they recorded (on videotape) pre- and posttreatment lessons for comparative purposes using a control group design. In both studies the questions 'broad', 'narrow', and 'evaluative' were analysed, and similar findings resulted: there was a significant increase in the number of 'broad' and 'evaluative' questions asked by teachers. However, even though the categories were limited in scope the two experiments show that by modelling behaviours in in-service settings, it is possible to get teachers to change their use of teaching skills.

In a contrasting experiment, the Ford Teaching Project at the Centre for Applied Research in Education, University of East Anglia, conducted ongoing in-service work based upon the principles of action research. Part of the project involved developing a pattern of in-service training in inquiry/discovery methods in a variety of subject areas. A number of hypotheses were developed as a basis for an operational definition of inquiry/discovery (Elliott and Adelman, 1974).

One focus of the Ford Project was teacher questioning. Fifty teachers were asked to engage in an ongoing re-examination both of

their own questioning practices as they related to the inquiry hypotheses mentioned above and of a number of hypotheses about questioning derived by the project team which they hoped teachers would examine in the light of their own questioning styles. Self-monitoring of lessons was carried out by teachers, and regular meetings were held at the schools to re-appraise questioning strategies as they related to the degree of communication between a teacher and his pupils. The meetings included small groups of teachers and research workers. The project included only those teachers who claimed to be interested in, or committed to, inquiry/discovery teaching (Elliott and Hurlin, 1974).

There was some indication that some teachers had difficulty in translating the ideals (hypotheses) into practice. It was reported that teachers found it difficult not to be subjective in their self-analysis. Another problem was that teachers had difficulty in developing a frankness with other teachers and researchers about what really happened in their classrooms in the lessons. Furthermore, many teachers felt threatened because of the constant change and complexity in their lessons as a result of continually analysing them (Elliott and Adelman, 1973).

To date, it has not been reported from this project whether the in-service work has resulted in improved teacher questioning, or improved pupil achievement. The research does, however, point to a need to achieve a balance in learning new skills, and for rapport between researcher and teachers.

To sum up the research on changing teacher questioning skills, then, there has been on the one hand a growing emphasis on identification of appropriate teacher behaviour, and on the other hand a concentrated effort to improve teacher competency in these behaviours. As Bush (1973, p.66) has pointed out:

> ... we need to be specific. Fewer global and so-called inspirational types of meetings (inservice) are needed. We should be precise in defining the teaching behaviour for which in-service educational programmes are designed. We ought to spend a much larger share of our total energy in helping teachers build an extensive repertory of technical skills.

This suggests that the search for the 'good' teacher, which has

dominated research on teacher effectiveness, should perhaps be replaced — or at least supplemented — by attempts to produce teachers who have a wide range of skills to apply in different situations in the classroom.

3.3 Teacher questioning behaviour and pupil achievement

It is not altogether surprising that, to date, so little has been accomplished in relating pupil learning to specific teacher behaviour. The number of relevant variables is so large that only a very 'tight' experimental programme would have any chance of identifying the factors involved, let alone their relative importance. Thus Rosenshine (1971, p.125) has had to conclude — after reviewing a number of studies in which teacher questioning had been related to pupil achievement — that no clear relationship was indicated between the frequency with which teachers used certain kinds of questions, and subsequent pupil performance. Even when researchers had experimentally induced increased use of particular types of questions, no consistently significant increases in pupil achievement had resulted. For example, Kleinman (in Rosenshine, p.125) found that while for high ability Grade 7 and 8 science pupils achievement was significantly correlated with the asking of more 'high level' questions, the results were not significant for the same teachers with their average and below average pupils.

There have also been problems when trying to compare the results of one study with another, because, as Rosenshine (1971) noted, there are frequently ambiguities in ground rules for recording observed behaviours, uncertainties in reported inter-rate reliabilities, and statistical inadequacies in data analysis. He also noted that there has been far less systematic observation of cognitive variables than of affective behaviours, and that the continued use of simplistic levels of recording, such as 'broad' and 'narrow' questions, has not helped in such little cognitive research which has been undertaken. Furthermore, analysis of results has been additionally complicated because most studies have used different classification systems. As a result, Rosenshine (1971, p.79) concluded that 'there are probably more differences between category systems in the methods for classi-

fying questions than in any other variable'.

Comparisons have also been made more hazardous because different tests have necessarily been used to measure pupil attainment following lessons with differing content and objectives. It is therefore not surprising that no clear pattern of relationship has yet emerged between the use of question-asking skills by teachers and their pupils' achievement. Some of the more significant studies will now be described.

In New Zealand, Wright and Nuthall (1970) identified several significant relationships between teacher behaviour variables and pupil achievement in seventeen Standard Two science lessons. Some teachers were found to ask several questions in rapid succession, and this habit was negatively related to pupil attainment (which was measured by a multiple-choice test of lesson content). A positive relationship ($r = .54$) was found, however, between asking just one question at a time and pupil scores. This probably reflected the fact that asking clear questions once is preferable — in terms of achievement — than repeating or muddling a number of questions into a single utterance.

In the same experiment, the frequency of asking closed and open questions was not significantly correlated with pupil achievement, but when data were converted to proportions instead of frequencies, the correlation between the use of closed questions and attainment was significant ($r = .46$), indicating that the greater the proportion of teacher questions of a closed type, the higher were the pupils' test scores. Another skill — redirecting of questions — was positively correlated with pupil attainment ($r = .54$), while two types of probing behaviours — extension and lifting questions — were not, for either frequencies or percentages. However, it is important to note that these findings were generated from the use of a test of mainly low-level knowledge. It did not, therefore, reflect as full a range of objectives as would be appropriate for inquiry-based teaching.

Complex patterns of relationship were found by Tisher (1970) in a study which <u>did</u> use tests which measured a range of thinking abilities. In secondary school science lessons taken by eight teachers, he found that moderate use of higher cognitive demands correlated directly with growth of understanding in science, for pupils of high ability, high prior knowledge, and high achievement orientation.

However, high- and low-use of higher cognitive demands by the teacher was associated with less growth of understanding for the same group of pupils. For pupils of low ability, low-level use of higher cognitive demands correlated directly with growth of science understanding, but use of higher-order questions was negatively correlated with achievement, as in the New Zealand study.

On a slightly different line of approach, Hunkins (1968) found that Grade 6 social studies pupils who worked daily for a month on sets of questions stressing evaluation and synthesis types, scored higher on evaluation and application test items than did other children who worked from sets of factual knowledge questions. Scores for both groups were the same for knowledge, comprehension, analysis, and synthesis items. However, other research has shown that it is extremely difficult to place elementary grade achievement test items, especially multiple choice items, clearly into Bloom's categories. The categories require very fine distinctions to be drawn in some cases (Klein, 1972). Also, there are reservations about the marking procedures used by Hunkins (Gall, 1970). In spite of these reservations, the Hunkins research indicates that it might be possible to improve children's ability to answer test items (questions) of particular kinds, through exposure to similar kinds of questions in lessons.

In a study by Ladd and Anderson (1970), teachers were divided into either those who practised 'low-inquiry' skills, or those who used more 'high-inquiry' skills in their teaching, on the basis of science discussions. The Smith and Meux categories were used to designate a question as 'low' or 'high'; low-inquiry involved operations such as defining, describing, and designating. Achievement tests were designed to reflect the two types of inquiry. It was found that pupils of high-inquiry teachers performed significantly better on tests which contained: (a) low-inquiry questions; (b) high-inquiry questions; and (c) a combination of the two forms of the test. Ladd and Anderson commented that further research would be needed to identify the processes involved in inquiry in specific contexts, such as classroom discussion, laboratory exercises, and reading assignments. However, Siddiqi(1973) noted that Ladd and Anderson did not report whether or not teachers actually did manage to teach lessons that could be categorised as low or high-inquiry. Nor was the actual

nature of the achievement test items reported.

It was indicated in Chapter 2 that creative thinking was often stated to be a goal of inquiry-teaching. Several researchers have reported that certain teaching practices have a positive effect on the creative thinking abilities of pupils (Getzels and Jackson, 1962; Torrance, 1962; Parnes, 1963). These writers point out that teachers who show tolerance, and give pupils chances to follow-up their ideas, will encourage creativity. It can be inferred from what these authors report that high-controlling, structured teachers will suppress pupils' creative expression.

From a study of teachers of Grades 2-5, Wodtke and Wallen (1965, p.80) found some support for the hypothesis that 'a high degree of controlling behaviour by the classroom teacher has a detrimental effect on verbal creativity, as measured by the Torrance tests' in Grade 4. Two high-controlling teachers, and two low-controlling teachers were found at each grade level by means of a Q-sort. Residual-gain scores were computed for pupil pretest and posttest creativity scores of fluency, flexibility, originality, elaboration, and total scores. The pretest to posttest delay was six months. Although no significant differences were found between pupils of each of the pairs of teachers at Grades 2 and 3, the Grade 4 pupils of low-controlling teachers scored significantly higher on verbal fluency, verbal flexibility and total verbal-creativity scores.

Yamamoto (1967) investigated whether more creative teachers made any difference in pupils' academic learning. No consistent results were found amongst nineteen Grade 5 teachers. High creativity among teachers did not usually result in better overall pupil achievement, nor in more favourable pupil adjustment.

Divergent thinking is a part of the Guilford Structure-of-Intellect model, and Guilford (1967) has claimed that divergent thinking could be improved — along with the other intellectual abilities in his model — through the use of appropriate teaching methods. Twenty-seven attempts to improve children's creative thinking ability were summarised by Torrance (1972). Tests of divergent thinking were the main criterion measures used and Torrance claimed that most of the studies showed improvements in children's creative thinking measured by this criterion. Dunkin and Biddle (1974) state,

however, that although it might seem 'self-evident' that classrooms high in divergent thinking will have better effects on pupils than those that stress convergent thinking, or rote memory, 'the evidence is simply not yet available' (p.255). They also state: 'evidence that teachers can be trained to seek divergent production is missing in research to date' (*ibid.*, p.256). There is a lack of agreement, therefore, about the effects teachers have upon creative and divergent thinking in pupils.

An experimental study conducted by Francis (1971), attempted to measure the effects of two kinds of science lessons upon pupil achievement. In one kind, the eight teachers in the sample taught a lesson aimed at getting pupils to learn and remember some new material. The same teachers taught another lesson with the intention of stimulating children to think. Results of three posttests of recall, recognition, and problem solving, indicated that when pupils were taught with the retention aim, their scores on the recall and recognition tests were higher than in the 'thinking' lessons. However, when results were corrected for the amount of content covered, there was no significant difference. Nor was there any significant difference between the classes aiming at retention and at thinking, on the problem solving test.

By using physics demonstration film loops, Suchman (1961) was able to provide Grade 6 pupils with guided practice in problem solving skills. Feedback and reinforcement were used by taperecording sessions, and getting the children to evaluate under teacher direction. It was found that after several weeks of such training, children asked many more questions judged by Suchman to be of better quality, their attitudes towards tackling difficult problems improved, and their self-confidence increased. However, there were considerable variations in the amount of change observed in different children.

In Grade 5 pupils of high intelligence, Suchman (1961) found a marked lack of autonomy and productivity in physics investigations, due, he believed, to the dependence of children upon authorities, teachers, parents, and books to shape their concepts. He thought that an associated reason was that the rewards in the classroom had come, typically, from giving the right answer. In such a climate, the generation of ideas seemed less likely to occur.

The results of the studies cited, indicate no clear, linear relationship between teachers' use of particular kinds of questions, and the achievement of their pupils.

Finally, one piece of research has looked at the effects of the situation organised by the teacher, rather than the teacher's behaviour. Crabtree (1967) studied creative aspects of Grade 2 pupils' play. In a crossover research design, each of two groups spent three weeks working on the basis of either a jointly determined structure of play, or a predetermined structure. From observations of pupils' actual play, divergent and convergent thinking, and constructiveness, were measured. Divergent thinking was said to be shown by children creating new patterns of play, and using materials in new ways. Convergent thinking was reflected in play which was conceptually accurate, logically deductive, and which led to correct conclusions. Constructiveness was exemplified in a continuum from routine to imaginative play. Observation scales were constructed for observers, and inter-observer reliabilities were high. Divergent thinking, and constructiveness in play, occurred more frequently in the jointly-determined programme. Convergent thinking occurred more often in the predetermined programme. However, the samples were small — only twelve children in each play group.

3.4 Related changes in teacher and pupil behaviour

Saadeh (1970, p.78) asked: 'Would anyone hire a so-called teacher who exhibits the teaching behaviour in class, but who has no effect on his pupils?' One of the goals of the present research was to find out whether changes in teacher behaviour, resulting from in-service treatments, can in turn lead to changes in pupil achievement as measured by achievement tests. More specifically, the focus of the research was to be on the kinds of questions teachers ask, but as already noted (Gall, 1970), there has been a paucity of research which examines the relationship between changes in the kinds of questions asked by teachers, and the achievement of their pupils.

Hutchinson (1967) found that, after training, four Grade 7 teachers used an increased proportion of higher level questions (these were convergent, evaluative, and divergent categories of the Aschner-

Gallagher interaction system). Their pupils showed significant growth on some of the creativity tests used, but there was no change on the other achievement tests constructed by Hutchinson. However, the study did reveal that when the teachers' use of high-level questions increased, the pupils' production of high level responses within the lessons also increased. It seems that when pupils sense this willingness on the part of the teacher to let them think, they might indeed begin to talk more freely and creatively. Thus it is possible that, under these circumstances, free response creativity scores might also improve. It is rather surprising to find that only one research study was located (that by Wallen *et al.*, 1969, which will be discussed shortly) which used free response creativity items in social studies achievement tests, in spite of the fact that inquiry teachers have been urged to employ open questions which give pupils a chance to express their own ideas.

A study by Flanders (1967) of Grade 7 and 8 teachers produced evidence that changes in teacher verbal behaviour induced by training can result in differences in pupil achievement. Lesson materials were provided for social studies and mathematics. Achievement tests were designed to test knowledge, and skills of problem solving, and the application of knowledge and skills relevant to unique or unusual problems. Teachers did not know the test items, but knew the tests' objectives. Teachers were taught the Flanders Interaction Analysis Category system by sound filmstrips. Observations of classroom interaction were made at the start, in the middle, and at the end, of subject units lasting two weeks.

The researcher found that the most indirect teachers asked longer, more extended questions, and did this about four times more frequently than did the most direct teachers. Ideas were dealt with in greater detail in the most indirect classrooms. The most indirect teachers did not interrupt the talk flow as often as the direct teachers, and made greater use of pupil talk. Pupil achievement was superior in the classes of the indirect teachers, at a level of significance beyond 0.01. Attitudinally, pupils favoured indirect teaching ($p < 0.01$). However, data for achievement tests were presented as a single score; subtests scores were not quoted. Therefore, it is not clear how the learning of facts was affected, compared with unique and unusual

problem solving items.

Snyder and Runkel (1975) reviewed twelve studies which reported attempts to change teachers' behaviour by training them to use the Flanders interaction system. Six of the studies reported 'second-step' effects, i.e., teacher changes resulting from training which resulted in pupil behavioural changes. These second-step changes were recognised by Snyder and Runkel as being particularly difficult to measure. Nevertheless, changes were observed in some studies after as little as nine to twenty hours of teacher training. The studies were limited, however, in the context of the present research, because pupil achievement was measured in each case by pupil talk, and not by prepared attainment tests.

Through in-service training of teachers of Grades 2-6 in social studies, Taba (1966) changed teacher and pupil behaviour in the following ways. Teachers decreased their use of 'seeks information' questions and pupils correspondingly decreased their amount of 'gives information'. An increase was found in teachers' use of 'seeks explanation' but there was no increase in pupils' use of 'gives explanation'. The category 'seeks generalization' increased in teachers' verbal behaviour, and so did the 'gives generalization' category for children. Taba *et al.* (1964, p.115) also classified discussion into high or low-level thought units, which relate to the degree of concrete or abstract thought involved in 'a remark or series of remarks which expresses a more or less complete idea, serves a specified function, and can be classified as a level of thought'. After training, teachers were found to have changed their patterns of question-asking and discussion sufficiently to decrease the number of pupils' concrete thought units and increase the number of highly abstract thought units. It was found that pupil use of high-level thought units was positively related to the teachers' use of units which sought to lift the level of thought.

The Taba research reported above did not relate teacher change to pupil achievement on test scores, but once again focussed upon students' verbal responses within lessons.

A large scale study which <u>did</u> test pupil attainment was that by Wallen *et al.* (1969). This research was an extension of the earlier work carried out by the Taba Project. Ten teachers who had worked with the Taba curriculum materials for at least two years, and who

were judged (by project staff) to be good teachers of the materials, were selected as an experimental group; a matched sample of teachers with no experience with the Taba curriculum was designated a control group. Intact Grade 6 classes were used, and the Curriculum and Control pupils were judged to be similar in abilities, as measured by a battery of tests, including an intelligence test, five tests specially constructed for the Taba Project, and the STEP test. Tests were administered in October of 1967 as pretests, and again (except intelligence) in May, 1968.

The teachers in the Curriculum group were given three one-day in-service courses, at intervals of one month, taken by the project staff. The two-fold focus was to help teachers develop content background for teaching cross-cultural topics (Latin America), and to help them learn strategies to be used in developing pupil thinking skills (which had been formulated by the Taba Project personnel). The in-service work on teaching strategies included the teachers studying and analysing tapescripts of the particular techniques to be learned and subsequently used in lessons.

It was hypothesised that the Curriculum group pupils would show greater achievement than Control group pupils, over the period of measurement. Analysis of covariance was used, with each pretest as a covariate. Because neither the teachers nor the pupils had been assigned by random means to the conditions, the statistical analyses were considered as suggestive only. In addition, because of administrative error, for some tests only six classes of each of the two groups were analysed.

The results showed that there were no significant posttest differences between the Curriculum group and Control group on the following tests: Ability to Explain; Application of Generalizations Test; and STEP. There were significant differences for the Interpretation of Data test (multiple choice); Generalizations — Simplified Score; Latin America — Human Score. This meant, in effect, that Curriculum group pupils showed greater gain in their ability to interpret data and make inferences, their level of concept development as applied to summarising stories, and concern for people-oriented social studies situations.

Greater gain was made by Control group pupils in ability to compare and contrast, utilizing concrete concepts, and grouping and labelling.

Consequently, no clear-cut superiority for the Curriculum group was found. Wallen *et al.* suggested that more than one year is needed for evaluating variables such as 'ability to explain' and 'application of generalizations'. A further conclusion of relevance to the present study, was that the Control group scored better than the Curriculum group at asking more conceptually abstract questions, suggesting to the project staff that the curriculum was limited in the extent to which pupils were able to raise questions and judge their worth. This in turn raises the question of whether or not the curriculum was too structured, and thus might not have allowed for the exploration of pupils' own ideas in a creative, inquiry mode.

There have been attempts by some researchers to induce changes in teacher questioning through means other than in-service courses, and then measure the effects of the changes upon pupil attainment. Using a complex experimental design, Church (1971) measured the effects of different kinds of questioning procedures upon pupils' test performance. Science lessons were taught to Standard 4 (ten year old) classes by Church himself in different ways which emphasised — in turn — these questioning techniques: highly structured closed questions; open questions with answers given by the teacher; open questions where pupils were not told the correct answers; open questions and prompting; and the use of infrequent questions. Of relevance to the present research is that Church found that if the length of the lesson was held constant, pupil achievement was higher when the teacher used highly structured questions and covered more lesson content than when only open questions were asked. When the teacher asked open questions and himself gave the answers after a short discussion, pupil achievement was lower than when the teacher required full discussion to elicit the answer.

Church's experiment is a significant contribution to the problem of how teacher behaviour affects pupil attainment both in terms of the verbal responses permitted of them and their performance on tests. His procedure of standardising both the lesson content and the actual teaching has been largely neglected in other research, to the extent

that Gall (1970, p.710) reported that he had found only two studies in the 1960s which had kept even lesson material constant.

A recent study which also standardised content and teaching method, was that by Franklin and Richards (1977), who tried to find whether children's divergent thinking abilities could be improved. Four classes of 9- and 10-year-old children were taught for one morning per week for six weeks by one of the researchers. All classes followed the same pattern, but two (one experimental and one control) were given formal exercises covering the same material as the other two classes (one experimental and one control), who did informal lessons. The divergent production exercises for the two experimental classes were aimed at getting children 'to generate information where the emphasis is upon variety and quantity of output related to a given source' (*ibid.*, p.68). The activities emphasised questioning skills and tried to get children to produce as many ideas as they could about a stimulus object. For example, in one session a brick was used to spark off answers to questions like: 'how can we describe it, what is it made of , what type of thing is it and what other kinds are there, where does it come from, what can happen to it, and what can be done with it?' (*ibid.*, p.68). Answers to questions like these were presented by the pupils in a number of ways, such as through drawings, writings, models, maps and graphs. The children were also encouraged to add new information and questions to the responses as the weeks of the experiment proceeded. Seven tests were used, five verbal, and two figural, from those developed by Wallach and Kogan (1965) and Torrance (1966). The experimental classes showed a significant increase on a number of these divergent thinking tests, compared with the control classes. These findings support the view that deliberate encouragement to pupils to talk freely, and ask questions about topics being discussed, can promote the development of divergent thinking abilities.

Also using the same teacher for all lessons, Ryan and Carlson (1973) tried to measure whether test performance is affected by two kinds of teaching: 'telling' and 'discovering'. The research design had an experimental group of children (Grade 1) who were given a chance to ask questions about, and discuss, lesson content they had listened to. A second group was given the listening experience without the follow-up discussion, and a control group was not given any treatment.

Analysis of variance revealed that the 'telling' group scored significantly higher than the 'discovery' group on an achievement test of the content. The researchers speculated that this might have been due to pupils' expectations — even at Grade 1 — that they should always give the right answer that the teacher wants, and that perhaps the 'discovery' pupils were confused because of the change of approach by the teacher. However, it should be noted that the test used was searching for 'right answers' anyway, and more could have been revealed by using tests that utilized the discussion and open-ended approach, as is the case with some tests of creativity.

In another approach Ryan (1973) tried to standardise lesson materials and teaching style, by supplying two teachers with prepared lesson scripts: one to emphasise the asking of recall questions; and the other to stress thinking abilities above the recall level. To match these scripts, two multiple-choice tests were devised by Ryan to measure 'low-level' achievement (recall), and 'high-level' cognitive achievement. Teachers were given lesson scripts which required them to ask a preponderance of either low-level or high-level questions in Grade 5 and 6 social studies lessons. The pupils of both these teachers achieved better scores than control group pupils on low-level achievement test items. On high-level attainment items, the high-level questioned inquiry group was superior to the low-level questioned group, which was similar to the control group. An implication to be tested by further research was that systematically-organised higher level questioning may encourage children to think at higher levels in inquiry-type social studies lessons.

In an extension of this study, Ryan (1974) used the same design (posttest only) with three groups of Grade 5 and 6 pupils. Ryan wanted to find whether getting teachers to extend pupil involvement by asking for three or four elaborations to the questions in the lesson scripts, instead of accepting one answer, made any difference to achievement test scores. Posttest scores showed that differences between the 'low-level' questioned class and the 'high-level' questioned class were not significant. Both these classes were significantly superior to the control group on both a retention test and a 'higher level' test. In Ryan's experimental design however, it seems that the control group did not work from the same lesson materials as the treatment groups. In that case, it would be surprising if the result could be

anything other than that the treatment groups would be superior!

3.5 Summary

It has been commented (for example, by Peck and Tucker, 1973; Borg, 1975) that in-service programmes to date have generally not been effective in changing the specific classroom behaviours of teachers. Perhaps the main reason is that the programmes have treated teaching behaviours in a generalised way?

In response to this assumed weakness a number of new approaches to in-service education have been developed over the last few years. In all of these there appear to be several identifiable principles which guide the way skills are taught. These are: clear identification of the skills; information about how the skills can be used; practice; feedback about performance; and chances to make modifications towards further improvement.

A good deal of evidence is now available which shows that the above techniques can successfully change teacher questioning behaviours (for example, Galassi *et al.*, 1974; Borg, 1975). A variety of question categories have been identified, usually in attempts to get teachers to employ a wider range of questions in their classrooms (such as divergent questions), but sometimes to reduce the use of other kinds (such as ambiguous questions).

It seems clear that in short in-service programmes (say up to 20 to 30 hours duration) only a few relatively straightforward skills can be taught. Complex skills such as those developed by Taba take much longer to master. In terms of teaching techniques, videotape is now frequently used as a means of teachers viewing their own teaching performance, but its use has no proven superiority over using written transcripts of questioning behaviour (Galassi *et al.*, 1974).

No clear relationship has been established between teacher questioning behaviour and pupil attainment from the few studies which have used attainment tests. Part of the problem is that different researchers have used different question categories and different kinds of tests of pupil attainment. Many of the tests have measured a narrow range of thinking.

Again, there is conflicting evidence about whether changes in teacher questioning that result from in-service programmes cause subsequent changes in their pupils' test attainment. Studies by Wallen *et al*. (1969) and Franklin and Richards (1977) have suggested that teachers' use of higher levels of questions can result in improvements in their pupils' attainment on tests which measure similar levels of thought.

Given that certain in-service programmes can change teachers' questioning behaviour, there is a need for researchers to try out various kinds of programmes related to particular educational objectives and instructional settings. Inquiry teaching in social studies is one such field. This would involve formulating question types which are appropriate for inquiry-based lessons because when used, they require pupils to use a wide range of thinking operations. Also, it would seem necessary for researchers to develop tests that measure similar thinking operations to find out whether pupil achievement is affected.

Some significant questions related to the issues raised above will be elaborated in the following chapter.

CHAPTER 4

DEVELOPING INQUIRY TEACHING: SOME HYPOTHESES

OVERVIEW: Arising out of the literature review presented in the previous two chapters, a number of unanswered questions related to inquiry teaching in social studies are identified. Particular attention is given to the following fields: teacher questioning; in-service training programmes designed to modify teacher behaviour; teacher attitudes; and the effects of teacher change on pupil attainment. A rationale for the study is outlined, and a number of research hypotheses relating to firstly, teacher variables and secondly, pupil attainment, are put forward.

4.1 Some unanswered questions

The review of literature above has indicated that, while there is no agreed definition of inquiry teaching, there are certain elements which can be identified. These elements are, first, that the teacher adopts a role in which he attempts to get children to engage in inquiry learning, a role involving a number of specific teaching skills. Second, when the pupils engage in inquiry learning, they need to employ a range of thinking processes to look at problem situations. Third, for these two elements to occur in the classroom there must be a particular kind of educational climate.

Numerous writers just cited have reported, however, that teachers usually dominate the classroom talk in social studies — as well as in other subjects — and in doing so, miss the chance to get children to think in varying ways. In much current classroom practice, there is a heavy emphasis upon low-level thought, in which memory recall is the dominant process, while little time is given to critical and creative thinking.

The question then arises, as to whether or not New Zealand teachers teach in a similar way in social studies lessons to those teachers observed overseas. For example:

1. What proportion of lesson talk is taken up by the teacher?
2. How frequently do teachers ask questions of various kinds in social studies lessons?
3. How often are questions asked that call for higher-level thinking on the part of pupils?

Assuming that the answers to these questions are similar to those found in other countries there is a need to explore ways of training practising teachers to use techniques that would involve pupils in discussion to a greater extent than is the case at present; to get children to talk more, to think more deeply and extensively, and to offer their own ideas and opinions. There is, however, very little evidence — in the New Zealand context — as to how this might be done effectively. It seemed worthwhile to explore various methods of in-service training with these purposes in mind.

The American microteaching studies cited in the literature review, have shown that an effective way to extend teacher skills (such as in asking questions), is to identify and define them, to give teachers a chance to practice them, and to provide feedback about how they were used. In-service courses that treat teaching skills in a general way (for example, by lecturing about them) do not appear to have had much effect upon teachers' actual behaviour.

This brings up a further series of questions:

4. Can teaching skills which help inquiry learning be operationally defined and communicated to teachers?
5. Is it possible, through in-service programmes, to influence teachers to change their proportion or frequency of use of such skills?
6. Are some kinds of in-service courses more effective than others in changing teachers' use of such skills?

It has been pointed out that the question-asking skills used by teachers are crucial in the inquiry process, but that teachers mainly ask the kinds of questions that require little more than memory recall and convergent thinking from pupils. It has been shown by a number of writers (for example, Gallagher, Aschner and Jenné, 1967) that although some 'factual' questions are necessary, too much emphasis on them by teachers uses up time which could have been spent getting children to expand their thinking.

While a number of observation systems have been devised for recording cognitive variables in classroom verbal behaviour, few have been developed for analysing teacher questions generally, and the sorts of questions that might promote inquiry in particular. Therefore, answers need also to be found to these questions:

7. What sorts of teacher questions are appropriate for inquiry learning in social studies?
8. Can teachers be taught to use the questions in their social studies discussions?
9. If teacher behaviour does change, do the changes last?

Teacher attitudes have been recognised as a significant aspect of inquiry teaching. Postman and Weingartner (1969) and Zevin (1969), for example, have argued that the teacher needs to have the sort of attitudes which allow him to be: flexible enough to allow pupils to express a wide range of viewpoints when discussing social issues; able to have children disagree with each other or the teacher; capable of letting children feel they are not constantly required to give only answers wanted by the teacher.

An additional question requiring further study then, is:

10. Can in-service courses improve teacher attitudes towards inquiry teaching?

From the research evidence of Borg *et al.* (1970) and others, there were good grounds for postulating that in the present study, an in-service course with a focus on specific question-asking skills, would result in modifications to teachers' posttreatment question asking. Few studies have, however, gone further, and tried to see whether alterations in teachers' questioning behaviour results in changes in pupil achievement. Literature related to teacher change and pupil attainment has indicated a need for tests which measure a wide range of thinking abilities in social studies. To date, however, most research has actually stressed a somewhat narrow range of thinking skills in the tests which have been employed. The ideal has not therefore been matched by practice. Little is known, too, about how long it would take for changes in teacher questioning (for example, the teacher asking more divergent thinking questions) to lead to improvement on test items measuring certain abilities (for example, pupils' divergent thinking skills) although there is some

evidence that suggests reasonably quick changes are possible (Guilford, 1967). Because most multiple-choice attainment tests of cognitive memory and higher-order thinking have a high correlation with intelligence test scores (Elley and Reid, 1969), it seems unlikely that teacher questioning would markedly affect scores on these sorts of achievement tests.

Some crucial questions here would be:

11. If a teacher asks more evaluative and divergent thinking questions, and fewer memory questions, do pupils' scores on memory items fall? Do they increase their evaluative and divergent thinking scores?

12. Does the teacher's use of the kinds of question which encourages pupils to infer and to be tentative lead to changes in scores on a test measuring inferential abilities?

It was questions such as those just asked which shaped the particular form of the present study.

4.2 Rationale for the research design

In planning the overall research programme it was therefore decided;

(a) to devise two alternative in-service courses on inquiry teaching — one would be a traditional information-giving course, while the other would emphasise specific questioning skills;

(b) to investigate the relative effectiveness of these courses in changing teacher behaviour in conducting inquiry lessons; and

(c) to see whether or not this led to any observed changes in pupil learning, as revealed by appropriate attainment measures.

Previous research, such as that by Gallagher (1965), has indicated that the Guilford Structure-of-Intellect Model (Guilford, 1956) provides a sound theoretical base for analysing classroom interaction. Aschner and Gallagher (Gallagher, 1965) used the operations categories of the Guilford model in constructing a complex interaction system to record cognitive variables. Included in this system were teacher questions of various types: cognitive memory (an amalgam of cognitive and memory operations), convergent, divergent

thinking, and evaluative questions. A fifth category, 'routine', was included to cover managerial-type questions.

These five categories were seen by the present investigator as being suitable for classifying initial teacher questions in social studies inquiry, because they covered a range of thinking operations. [Smith and Meux (1962) identified initial questions as those which begin a verbal exchange between two or more people.] In addition, the five Aschner-Gallagher types could be defined (with some slight modifications explained in Chapter 6) in terms that could be readily communicated to teachers in in-service programmes.

However, it was also considered important to look beyond the initial questions asked by teachers to the sort of questions that required pupils to think more about a response they had already given. In this respect, two types of 'follow-up' questions were identified as important for social studies inquiry: 'extension' and 'grounding'. Both kinds usually depend on a previous question, and because teachers cannot always anticipate when they will occur, or their precise form, there appeared to be a need for in-service settings for practice in asking them. The necessity for extension-type questions to be included in analyses of questioning has been mentioned by numerous writers (for example, Taba *et al.* 1971; Gall, 1970; Nuthall, 1970, p.17) particularly because they allow a teacher to, firstly, get feedback from initial questions, and then to ask additional questions to develop thinking further. Grounding questions are a form of probing which requires a pupil to give reasons or evidence for an initial answer he has given. The importance of this kind of question has been recognised by, for example, Borg *et al.* (1970), Turney *et al.* (1973) and Massialas and Sprague (1974). Grounding is of particular significance to social studies discussion since, as Massialas and Sprague have pointed out, what distinguishes an inquiry discussion from a non-inquiry discussion is that in the former the children are required to ground (i.e., give evidence for) the answers they give.

The seven types of questions identified above thus formed a system which could not only be readily explained to practising teachers but could also provide a considerable amount of systematic information about the questioning patterns employed in typical social

studies lessons.

It was then decided to design the main experimental in-service programme in such a way that teachers would deliberately set out to practise the seven kinds of questions; this would encourage them to explore the whole range of question types, using all of them rather than over-emphasising a few. If this goal were achieved it was expected that teachers would broaden their repertoire of question asking skills, and that in posttreatment lessons, they would decrease their dependence on cognitive memory, convergent, and routine questions, and increase their use of evaluative, divergent thinking, grounding, and extension questions.

To find out whether this in-service skill-based programme had had any overall effect on the pattern of teacher question-asking, it was also decided to devise an inquiry ratio (or score) based on the formula:

$$\frac{E + D + Ex + G}{T}$$

where E = number of evaluative questions.
D = number of divergent thinking questions.
Ex = number of extension questions.
G = number of grounding questions.
and T = total number of questions.

It was postulated that, if the in-service training were successful, then observations of the teachers concerned would produce a higher inquiry ratio (i.e., use a larger proportion of types of questions conducive to inquiry learning) in posttreatment lessons than they did in pretreatment lessons.

Other aspects of teacher behaviour were also of interest. For example, research has indicated a need to increase pupil participation in inquiry discussion, without undue intervention and domination from the teacher (Massialas and Cox, 1966; Bellack *et al.*, 1966). That is, teachers should let talk flow in extended patterns of interaction to reduce the dominance of the "teacher seeks, pupil responds, teacher reacts" sequence, which Bellack had identified as the predominant one in most classrooms. It was therefore decided to find out also whether posttreatment lessons differed from

pretreatment ones in the frequency of occurrence of the following interaction patterns:

1. Teacher talks - pupil talks - teacher talks.
2. Teacher talks - two or three pupils talk - teacher talks.
3. Teacher talks - four or more pupils talk - teacher talks.

Furthermore, because the attitudes of teachers were considered important in inquiry teaching, it was considered desirable to find out whether in-service education about inquiry caused any attitudinal changes. And to seek answers to the questions about teacher change and pupil achievement raised in the previous section, it was decided to construct achievement tests for pre- and posttreatment lessons which would be subdivided into four groups of items to measure: cognitive memory, convergent, evaluative, and divergent thinking abilities. These would be supplemented by pre- and postmeasures of inferential thinking. These tests are fully described in Chapter 6.

4.3 Hypotheses related primarily to teacher behaviour

In order to answer questions such as those set out earlier, three in-service training treatment groups were thus proposed:

(a) an Experimental Skills (ES) group, whose programme would concentrate on practise of specific questioning skills;

(b) an Experimental General (EG) group, whose programme would deal in a more general way with advantages and requirements of inquiry teaching; and

(c) a Control (C) group, who would not take part in either type of programme.

It was hypothesised that, as a result of the in-service course treatments, teachers in the Experimental Skills group would modify their questioning behaviours to a greater extent than would teachers in the Experimental General group, and that both of these groups of teachers would change more than those in the Control group.

Stated in general terms, the hypotheses to be tested can be grouped as follows:

H1 As a result of their various in-service programme treatments ES group teachers will:

1.1 increase their actual frequency of asking Evaluative thinking, Divergent thinking, Grounding and Extension questions and decrease their actual frequency of asking Cognitive memory, Convergent thinking and Routine questions;

1.2 increase their proportionate frequency of asking Evaluative thinking, Divergent thinking, Grounding and Extension questions and decrease their proportionate frequency of asking Cognitive memory, Convergent thinking and Routine questions;

1.3 decrease their proportion of teacher talk;

1.4 increase their Inquiry Score;

1.5 increase their use of extended patterns of Teacher-pupil interaction

in relation to teachers of the EG and C groups, in a discussion lesson situation immediately (i.e., within two weeks) after the in-service programmes.

H2 The same relationship between the groups in teacher questioning behaviour will be observed in a discussion lesson situation three months later.

H3 Teachers who have taken part in in-service courses on inquiry teaching will exhibit a more positive attitude towards inquiry teaching, and less dogmatic attitudes than teachers who have not taken part in such courses.

4.4 Hypotheses related primarily to pupil behaviour

In addition to the above hypotheses a number of hypotheses were formulated which related primarily to pupil behaviour. These can be stated as follows:

H4 As a result of the various in-service treatments of their teachers, pupils of ES group teachers will:

4.1 not change in their attainment in the Cognitive memory subtest;

4.2 not change in their attainment in the Higher-order thinking subtest;

4.3 increase their attainment in the Evaluative thinking subtest;

4.4 increase their attainment in the Divergent thinking subtest,

in relation to pupils of the EG and C group teachers, in tests taken immediately (i.e., within two weeks) after the in-service programmes.

H5 The same relationship between the groups in pupil attainment will be observed in the various subtests three months later.

H6 Pupils of the ES group teachers will increase their posttest attainment on an inference test in relation to pupils of EG and C group teachers.

These hypotheses are restated in a specific form appropriate to statistical testing in Chapter 7.

CHAPTER 5

DESIGN OF THE STUDY

OVERVIEW: The experimental design of this study is described. Twenty-four teachers of Form One and Two classes were assigned to three treatment groups in a non-equivalent control group design, with the usual class of each teacher making up the pupil sample. A rationale for the selection of the sample is presented, and a description of the teachers and pupils given, along with procedures of selection. Data collection methods are briefly described, as are the statistical analyses of the data relating to teacher and pupil variables. An outline is given of each of the two in-service programmes which formed the experimental treatments: the SPRITE course which focussed upon specific questioning skills, and a more general course on inquiry teaching in social studies. Each teacher taught three lessons based on standardised substantive content developed by the investigator, and presented to pupils by means of television. The method of construction of these materials is described.

5.1 Experimental design

It was stated in Chapter 4 that the first major aim of this study was to find out if a skills-based in-service programme (on teacher questioning in inquiry teaching) would result in changes in the verbal behaviour of the teachers concerned. Accordingly it was necessary to observe these teachers in a standardised situation, both before and after their in-service treatment. As the primary interest was in the kinds of questions they asked, it was also necessary to taperecord their lessons so that transcripts could be made and subsequently analysed.

It was decided that observation lessons would be taperecorded as follows:

1. Each teacher would take one lesson, using standardised material, with his own class, within three weeks prior to the experimental treatment.

2. Another lesson would be taken in each classroom within two weeks following the experimental treatment.

3. A third lesson would be taught by each teacher twelve weeks after the previous one.

To determine whether any behavioural changes in the skills group — from pre- to posttreatment — could be attributed to the experimental treatment and not to other causes such as the influence of practice, the following experimental design was established:

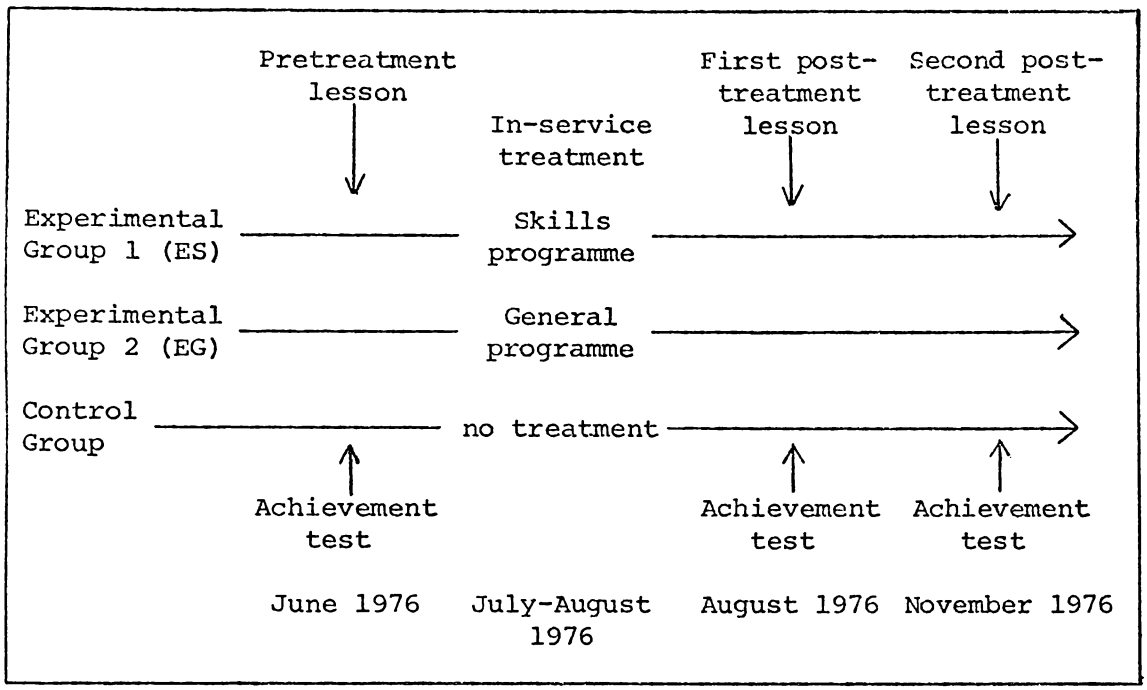
1. Teachers in the Experimental Skills (ES) Group would be given an in-service programme (treatment) which focussed upon skills of questioning relating to inquiry teaching in social studies.

2. Teachers in the Experimental General (EG) Group would be given an in-service programme which focussed only upon general aspects of inquiry teaching in social studies.

3. Teachers in the Control (C) Group would be the subject of exactly the same observational procedures as those in the two experimental groups, but would not receive any in-service training.

The second major aim of this study was to find out whether changes in the verbal behaviour of teachers resulted in changes in the achievement test scores of their pupils. Achievement tests, based upon the substantive content of each of the three observation lessons, were therefore constructed and were administered by the experimenter to each class on the day following the lesson concerned.

The experimental design can be summarised in the following diagrammatic way:



The design, then, was based on the 'nonequivalent control group design' described by Campbell and Stanley (1963, p.217). Each of the classes constituted 'naturally assembled collectives ... as similar as availability permits but yet not so similar that one can dispense with the pretest' (*ibid.*).

5.2 Description of the sample

5.2.1 Rationale for the research sample

Teachers of Form One and Two classes (children's ages approximately 11 to 13 years) were selected for this study because they represented part of the Form One to Four level covered by the *Draft Syllabus in Social Studies* (1972). This syllabus attempted to focus upon inquiry in social studies, including teacher questioning behaviour. Therefore, it was considered that research findings in this area would be a useful addition to the small amount of available information at present on the effects of this type of curriculum. Also, the pupils were considered to be of a suitable age to be tested, through achievement measures, on a variety of thinking abilities. Piaget's investigations have suggested that it is not until a child is at least eleven years old that his thought processes are capable

of handling abstract problems, and that he is able to hypothesise beyond the cultural problems and situations which are immediately familiar to him (Hilgard, 1967). To provide as wide a range of information as possible, the lesson materials and tests were designed to get the learners to think abstractly, as well as concretely.

Because of the extensive observations and analyses required, the study had to be restricted to a comparatively small number of teachers. Accordingly it was decided to confine it to teachers of one sex only, to avoid a possible confounding variable. Ryans (1960) found that there were some significant differences in the traits of male and female teachers in the fields of interests and activities. He recognised these differences as potential sources of bias in the direction of the more numerous sex in a research sample. For example, it was found that elementary school men teachers scored significantly lower than elementary school women teachers on the characteristic representing responsible, systematic, and businesslike behaviour. Men were found to be more favourable, in attitude, towards democratic pupil practices, and more inclined towards permissive, child-centred viewpoints, and more emotionally stable than women teachers. Teachers of social studies in secondary schools reflected the same trends; in addition, men scored higher than women on 'emotional adjustment' (*ibid.*, p.296).

Differences between male and female teachers were also reported by Adams and Biddle (1970), who found that in male teachers' classrooms (in social studies and mathematics), a greater amount of information dissemination took place than in female teachers' classrooms. Women teachers spent more time 'intellectualizing' about organisational matters. Also, male teachers' classrooms tended to be more centrally organised, which was associated with the teacher in a central group, audience-only situation, with the teacher as the emitter (that is, the person who speaks first in verbal interaction). Their data showed, however, that more 'disengagement' took place in male classrooms: that is, in some classroom interactions, some pupils took no part whatsoever. The authors concluded that the sex of the teacher has significant effects upon the teacher-learning situation.

The present study, therefore, was confined to male teachers.

Although the generalizability of its findings to the teaching population as a whole was reduced by so limiting the sample, more effective matching within and between groups was thereby achieved.

5.2.2 Description of the teacher sample

Twenty-four teachers were included in the research sample, enabling eight teachers to be allocated to each treatment group. This number was the maximum from which the experimenter could collect observational data within a reasonable time-span, and groups of eight were considered to be large enough for comparative purposes if analyses of variance or covariance were employed.

The intact classes of these teachers formed the pupil sample. For data analysis, 523 children remained who had completed all lessons and tests, out of a total of 794 in all classes at the start of data collecting. There was no evidence of differential mortality in the pupil experimental groups, for a check showed that in each class, approximately the same number of girls as boys 'dropped out' (mostly due to illness), and these pupils covered more or less the full range of abilities as compared with the remainder of the class.

The twenty-four teachers and their classes were located in all five Intermediate Schools in Hamilton City, and one Intermediate School located in a 'satellite' town, twenty kilometres from Hamilton. Hamilton City has a population of approximately 100,000 people and the population of the satellite town is about 7,000. It was necessary to select teachers from one school outside Hamilton City in order to achieve a numerical balance in the treatment groups and control group.

Four teachers were selected from each school, as this was the maximum number any school Principal could release for the in-service experimental treatment at the same time. Therefore, to avoid possible contamination effects between treatment groups, the four teachers in each school were assigned to the same treatment group. Thus, the eight teachers in each treatment group were from the same two schools.

The teacher sample was selected in the following way. The investigator met with the Principals of the five city schools two months prior to the data collection phase. The purpose of the

research was outlined in broad terms, and the Principals accepted that precise details might lead them (Principals) inadvertently to 'tell teachers too much' when discussing the project. The investigator was known to all Principals, and this factor assisted in gaining a positive response. Problems of whether the schools could release four teachers at the same time for an in-service course, suitable times for classroom testing and recording, and best ways to contact teachers, were all discussed. All Principals agreed that the problems could be overcome. The same procedure was followed in a separate meeting with the Principal of the satellite-town school.

The investigator visited each school, and outlined the project at a meeting of male teachers. The outline was in broad detail, so that teachers did not know the specific purposes and research design regarding treatment groups. Volunteers were called for, and each volunteer was asked to complete a form giving details of age, years of service, class taught, and whether he had attended any in-service courses in social studies within the previous two years.

In the six schools concerned, there were 44 male classroom teachers available for such a project, 36 of whom volunteered to take part. The criteria used in selecting the sample were that the teachers were male, taught a relatively heterogeneous class in terms of abilities, and had not participated in a social studies in-service course of a duration of one day or more, within the two years prior to June, 1976.

From this pool of teachers, three groups of eight were formed. Teachers were matched on the basis of age, years of service, and class taught (Form One or Form Two). Schools were randomly allocated to treatment or control groups. When more than one teacher from one school was available for any category, selection was made by random means. No attempt was made to get 'good' teachers. It was hoped that each group of eight would be heterogeneous in teaching abilities and styles. Table 5.1, p.185 shows details of the matching of groups and it can be seen that the three groups were very similar on the basis of the variables used.

The average age of the ES group teachers was 33 years (range 25-45 years), of the EG group 33 years (range 25-45) and the C group 34 years (range 25-45 years). The mean years of teaching experience

was 9.8 years for both Experimental groups, and 11.4 years for the Control group.

Of the 24 teachers who were included in the sample, all were still available for the first posttreatment lesson, but two were unavailable for the second posttreatment recording (one teacher from the ES and one from the C group). These were unavoidable drop-outs, because both had taken up teaching positions elsewhere during the intervening period.

5.2.3 Pupil sample

It was intended to select teachers who taught heterogeneous classes; that is, each class would contain pupils of a wide range of abilities. Intelligence quotients, which were available for each class, indicated that twenty-three of the classes were similar in distribution of abilities. One class, in the ES group, was found to include a disproportionate number of above-average ability pupils. However, through the use of analysis of covariance, initial inequalities were compensated for when determining pupil change. The age range of the pupil sample was 10 years 1 month, to 13 years 6 months, as at 1 January 1976. The number of pupils who took part in all lessons and tests in each class is shown in Table 5.2, along with their mean age and mean intelligence quotient.

It was assumed that the classes within each treatment group, and between groups, were 'approximately similar' for the purpose of internal validity. The design was regarded as having controlled the main effects of history, maturation, testing, and instrumentation. As Campbell and Stanley (1963) comment, the use of intact classrooms for experimentation is less likely to create over-sensitivity to the experiment in pupils; the 'I'm a guinea-pig' attitude.

It was also considered that there were insufficient socio-economic differences in the catchment areas of the six schools in the sample, to unduly bias pupil samples. Each school was attended by pupils from a wide range of socio-economic backgrounds, including the satellite town school. This town has a close connection with Hamilton City because of substantial numbers of people who commute to the city each day for work. A small proportion of pupils of all schools is made up of rural children, although the proportion was highest for the satellite-town school. Ethnically, the distribution of pupils was

similar in all schools.

5.3 Procedures

5.3.1 Collection of data

The experiment was carried out during Terms Two and Three of 1976 (June to December). Full details of the procedures are presented in Chapter 6, but briefly, the data were collected as follows:

(i) Prior measures of pupils' scores on the *Progressive Achievement Test (PAT) Reading Comprehension* (Elley and Reid, 1969), and the *Otis Intermediate Intelligence Test* were gathered, along with pupil ages, to be used as possible covariates in the statistical analysis procedures to measure pupil change in achievement.

(ii) The investigator and an assistant administered the *Social Studies Inference Test* (McNaughton, 1974) to the 24 classes in the sample, in June and again in November-December.

(iii) In July, each of the 24 teachers taught a twenty-five minute lesson based on objectives and content supplied by the investigator, who also taperecorded every lesson.

(iv) Following the first of the taperecorded lessons, teachers in the Experimental Skills Group and Experimental General Group attended separate four-day in-service courses conducted by the investigator. The control group was given no treatment.

(v) On the day following each lesson, an achievement test was administered to the pupils by the investigator or an assistant.

(vi) Two further lessons were recorded in each classroom: one within two weeks after the experimental groups had received their in-service 'treatment' (August) and the other after a delay of twelve weeks (October-November).

(vii) Teachers also completed the Zevin *Attitude-Towards-Inquiry Scale* and the *Rokeach Dogmatism Scale, Form E* before the

experimental treatment and again after the final tape-recorded lesson had been taught.

(viii) The taperecordings of the lessons were transcribed, and from the transcriptions, trained coders classified the questions asked by teachers. All tests that had been administered were marked by the investigator.

5.3.2 Analysis of data

The research design involved the analysis of data for two major groups of variables: those relating to teacher behaviours and those relating to pupil behaviours. Teacher variables were measured to find whether or not there were any posttreatment differences among teachers which might be attributed to treatment effects, and the criterion variables were: types of teacher questions, inquiry score; patterns of interaction; and teacher attitudes.

Data for teacher variables were analysed by means of the B6700 (Auckland University) computer using programme 'Teddybear' (J.B. Wilson, University of Otago, New Zealand). The programme was used to compute multivariate analysis of variance for pretest variables grouped as follows: proportion of each type of question asked; frequency of each type of question asked; proportion of various interaction patterns and proportion of teacher-to-pupil talk. It was assumed that if no significant differences were found in the pre-treatment data, then one-way analysis of variance would be appropriate for each dependent variable for pretreatment and posttreatment data. If significant differences were found as a result of the multivariate analysis of variance, then computations would be extended to determine which dependent variables contributed most to the significance of the interaction (Cooley and Lohnes, 1971).

Analysis of variance was considered appropriate because the samples were drawn from an entire male teacher population of six Intermediate schools, and variance within the three groups would be assumed to be approximately equal, prior to experimental treatment. The basic principle of analysis of variance 'is to determine whether the sample means vary further from the population mean than we should expect, in view of the variation of single cases from their means' (Guilford and Fruchter, 1973, p.230). The technique involves two

estimates of population variances: between-sets sums of squares and mean square; and within-sets sums of squares and mean square.

It was planned to compute F ratios for each one-way analysis of variance computed for each teacher variable. If significant differences were found, Duncan's *New Multiple Range Test* (Huck, Cormier and Bounds, 1974) would be used to determine which of the three groups were significantly different from the others at either the $p \leq .01$ or $p \leq .05$ levels of significance.

In this study, the $p \leq .05$ level of significance, a figure widely accepted in educational research of this type (Guilford and Fruchter, 1973), was selected as the level for rejecting the null hypotheses.

Pupil variables were measured to find out whether or not there were changes in the posttreatment test scores of pupil groups which might have resulted from changes in teacher behaviour. The criterion variables were pupil scores on achievement subtests together with inference test scores. Because intact classes were used, the analysis of covariance method was selected to analyse pupil variables. Winer (1971) points out that this procedure is a robust technique for a situation where natural settings are used as the basic sample units. Campbell and Stanley (1963) recommended analysis of covariance as appropriate for a pretest-posttest, control group design using natural classes as the units of analysis.

The question as to which is the best method of measuring pupil change has been extensively argued. The use of 'raw' gain scores has been criticised as far too simplistic, and likely to produce misleading results (Cronbach and Furby, 1970). Analysis of covariance and residual gain scores are the two most favoured techniques, and Linn and Slinde (1977) have claimed that no clear superiority had been established for either method. Lord (1967) cautioned that there were difficulties in making proper adjustments for any pre-existing differences between groups. Whereas the residual gain score is a way of singling out individuals who change more (or less) than expected, analysis of covariance uses regression analysis to sort out group differences, holding other factors constant. Since group differences were the primary interest in this study, analysis of covariance was employed.

Data for potential covariates were collected before the main data collection phase began. These were quotients from the *Otis Intelligence Test; PAT Reading Comprehension* scores; and age. These data were to be used with pretest achievement scores, in stepwise regression analyses, to determine the most appropriate covariates. To compute significance levels of differences between groups, Duncan's *New Multiple Range Test* was again used, with significance levels set at $p \leq .01$ and $p \leq .05$ as before. The analysis of covariance was calculated using the computer programme 'Teddybear' (J.B. Wilson).

For inference test data, simple analysis of variance was used to test for significant differences between groups on their pretest scores, and again on their posttest scores.

5.4 The in-service programmes

Two four-day in-service courses were conducted by the investigator: one for the Experimental Skills (ES) Group teachers, and one for Experimental General (EG) Group teachers. There were eight teachers in each treatment. The courses were held in a geography seminar room at Hamilton Teachers' College which meant that a valuable resource base for social studies materials relevant to Intermediate schools was readily available.

Each in-service course consisted of twenty-five hours of instruction-workshop activities. The ES Group teachers (N = 8) attended on Thursday and Friday of two successive weeks. The EG Group teachers (N = 8), attended on Monday and Tuesday of successive weeks. The order was decided by chance. The several days between each two-day attendance provided an opportunity for teachers to read related materials, and practice suggested procedures in their usual classrooms. Each in-service course day was divided into three work sessions: 8.30 a.m. - 10.30 a.m.; 10.45 a.m. - 12.30 p.m.; 1.30 p.m. - 4.00 p.m.

During each course, the investigator compiled anecdotal records of teachers' reactions to the programme and the teachers completed an evaluation sheet at its conclusion. At the start of each course, the two attitude scales (Rokeach Dogmatism, Form E and the Zevin scale) were administered.

5.4.1 Objectives of the courses

The attention of ES Group teachers was focussed upon specific inquiry discussion behaviours, including types of questions for teachers. The other course focussed upon inquiry teaching for social studies in a more general way, but exactly the same material relating to inquiry skills was introduced in both courses.

5.4.2 In-service programme for the ES Group: SPRITE

It was indicated in Chapter 3 that there appear to be a number of essential elements which have to be built-in to in-service courses if they are successfully to alter teachers' classroom behaviour. Taking these elements into account, a particular form of in-service programme was developed for this research: Sensitization, Practice, and Review, In-service Teacher Experience (SPRITE). The SPRITE programme was based upon the following assumptions:

(i) That inquiry teaching is effective for teaching cross cultural topics in the 'new' Form One to Four Social Studies Syllabus for New Zealand.

(ii) That skilful teacher questioning facilitates the process of inquiry in class discussion.

(iii) That teachers spend considerable proportions of class discussion time asking questions, and therefore teachers' skill in asking questions is an important concern for in-service education.

(iv) That short, concentrated courses like SPRITE are likely to be more effective in producing changes in teachers' verbal classroom behaviour if they:

a. Focus upon specific skills for teachers.

b. Sensitize teachers to the specific skills, by providing operational definitions, and examples.

c. Provide for practice of the skills in workshop and classroom lessons.

d. Review teachers' practice performance and help them modify it for future lessons.

(v) That the in-service course (SPRITE) has a sound theoretical basis — a necessary condition which was stressed by Taba

(1966) and outlined in Chapter 3 of this study.

The first day of the SPRITE programme proceeded as follows:

(i) The Rokeach and Zevin attitude scales were administered.

(ii) A lecture, 'Recent developments in social studies', was given. The topics included developments related to content, thought processes, inquiry teaching, values and attitudes, resources and discussion techniques.

(iii) Through 'brainstorming' and reporting back, followed by the issue of notes from Zevin (1969), characteristics of 'inquiry' teachers in social studies were identified.

(iv) A lecture-discussion was conducted on 'open -ended teaching'. Values education topics were used as case study material.

(v) Using materials supplied by the investigator, teachers planned lesson sequences designed to reflect inquiry teaching. Each teacher 'taught' his lesson to the other teachers, and attempted a rationale for this particular inquiry-type of lesson.

(vi) The teachers were asked to read Chapter Three of Postman and Weingartner (1969): *The Inquiry Method.*

(vii) Teachers were shown a substantial collection of books and periodicals assembled in the seminar room (Appendix A). Arrangements were made for teachers to borrow materials.

The following content and activities made up the second day:

(i) A two-hour discussion was conducted, based on Postman and Weingartner, Chapter Three (1969). A reading of Cordier (1968) and a lesson transcript (Massialas and Zevin, 1967) were also used in discussion.

(ii) The teachers were sensitized to teacher question types by working systematically through a reading *Teacher Questions in Social Studies.* This was written by the investigator, detailed the operational definitions, and gave examples of each primary question type in the classification system outlined in Chapter 6. The reading is reproduced in Appendix B (p.260).

(iii) Teachers planned lessons based on cross-cultural materials. The teachers were asked to build into their lessons a wide range of teacher questions. Arrangements were made for the teachers to taperecord their ten-minute lessons with a group of children in the three school days before they returned for the third day of the course.

On the third day, the teachers were engaged in these activities:

(i) The lessons taught by the teachers were reviewed. Each tape was played, and all teachers attempted to classify the questions asked into one of the four primary types. The amount of teacher talk was also recorded. Discussion was held to offer feedback to each teacher about whether or not more variation could have been built into the lesson. Divergent questions were studied at length, for these proved difficult for teachers to frame. This phase took most of this day's work.

(ii) The teachers were sensitized to 'grounding' and 'extension' questions in the same way as for primary questions. A reading of Massialas and Sprague (1974) was used to identify inquiry sequences using these types of questions.

The last day of the SPRITE programme was spent in this way:

(i) A range of social studies materials was discussed. These materials dealt with cross-cultural topics, and were considered suitable — by the investigator — for inquiry teaching. These were the materials also used by teachers of the Experimental General Group.

(ii) A lecture was given on children's thinking in social studies, based upon Michaelis (1975). Teachers practised formulating questions to foster inferential and tentative thinking.

(iii) After viewing a film excerpt about an African community, the teachers were given further practice in writing lesson sequences, which utilized all question types in the classification system. The investigator demonstrated how memory and convergent questions could often be replaced or reduced by using evaluative or divergent thinking questions.

(iv) A lecture was given to summarise the main emphases of the SPRITE course: a rationale for inquiry, characteristics of inquiry teachers and pupils, and question types. It was stressed that the course hoped to offer specific behaviours that teachers might find useful to experiment with in their teaching of social studies.

(v) An evaluation form was completed.

(vi) The lesson material for the posttreatment lesson was shown to the teachers, and arrangements were made for recording the first posttreatment lesson during the following fortnight.

5.4.3 In-service programme for the EG teachers

There were two main differences between the course for Experimental Group teachers, and the SPRITE in-service programme. First, the content about teacher questioning skills was treated in a more general way for EG teachers. Second, more time was spent working on the development of teaching materials and lesson sequences, using cross-cultural materials. The 'general' course was based on the following assumptions:

(i) That teachers' classroom behaviour can be changed by a general course of lectures and workshops on inquiry teaching in social studies.

(ii) That the possibility of such changes will be enhanced by studying appropriate cross-cultural materials and the planning of actual lesson sequences.

The EG programme took the following form:

(i) The Rokeach and Zevin attitude scales were administered.

(ii) The lecture given to SPRITE teachers was also given to these teachers 'Recent developments in social studies'.

(iii) A lecture was given by the investigator on 'The characteristics of inquiry teachers and pupils'. The content was the same as that used in the SPRITE course, but the lecture was more formal. Readings and transcripts of social studies lessons were issued as recommended reading, and teachers were left to do follow-up reading independently.

(iv) A second lecture, 'Values education', was given in which recent materials in values education were outlined and shown and consideration was given to the role of the teacher as a promoter of open-ended discussion.

(v) Slide-tape teaching sequences were demonstrated by the investigator, and these stressed the use of inquiry teaching and a variety of questioning procedures. The teachers searched for materials — from the available library resources — suitable for their classes.

(vi) The same social studies references (Appendix A) shown to SPRITE teachers were also held in the room for these teachers, and arrangements were made for borrowing.

On the second day, the teachers were reminded of the readings relating to inquiry methods in social studies. The readings were available for the teachers to take away (Zevin; Cordier; Postman and Weingartner; lesson transcripts; *Teacher Questions in Social Studies*).

The EG teachers spent most of this second day developing teaching materials. These focussed upon audio-visual materials, and the topics were designed to fit into the Form One to Four Syllabus (Draft), so that they would be suitable for the teachers' own classes. Arrangements were then made for teachers to take lessons before the next day of the course, in which they would try out the lesson materials they had prepared.

The third day was conducted in the following way:

(i) Each teacher showed the lesson material which had been prepared on Day Two, and which had been taught to his class. This took over half the day. After each demonstration, the materials were discussed, in relation to how useful they were in teaching cross-cultural topics.

(ii) A lecture called 'Games to promote inquiry in social studies' was given, and this was followed by experience in playing simulation games, and in discussing their usefulness in getting children to explore ideas.

(iii) Lesson sequences developed during the second day were completed.

On the fourth day the following areas were covered.

(i) A lecture, 'The use of other audio and visual aids in social studies', was given. This included a study of pictorial materials and overhead projectors, and how they could improve learning.

(ii) In the workshop situation, the teachers prepared further audio-visual lesson materials for their classes. These included tape-slide sequences, written teaching, simulation games, and overhead transparencies.

(iii) The teachers completed the course evaluation form, and the lesson materials for the posttreatment lesson were shown to the teachers. Following this, arrangements were made for recording the posttreatment lesson during the following two weeks.

5.4.4 Summary

The two in-service courses both dealt with inquiry teaching in social studies. But whereas the SPRITE treatment focussed upon sensitizing, practising, and reviewing specific question-asking skills, the second treatment was considerably more general. In the SPRITE programme, teachers were engaged in intensive study of skills which had been operationally defined by the investigator. The practice of the skills was carried out in the teachers' classrooms, and subsequently reviewed. This work was supported by literature and lectures relevant to inquiry teaching and to specific questioning skills. The teachers in the more general course were given the same literature, but were left to read it independently. In the subsequent analysis of posttreatment lessons, it was hoped to find whether these parallel but differing treatments led to differences in teacher questioning behaviour.

5.5 Lesson materials for recorded lessons

The materials for the three lessons taught by sample teachers, and taperecorded for analysis, were developed by the investigator. All materials were tried out in a school well removed from the schools to be used in the main experiment. Two pilot teachers used the

materials, and after discussion for feedback about their effectiveness, modifications were made.

By standardising the materials for teachers and pupils, extraneous factors associated with content variation were reduced substantially. Dunkin and Biddle (1974) have noted that failure to standardise lesson content made judgments about comparability between teachers dubious. The materials used were assumed to provide each teacher with the same opportunity in conducting a discussion with his pupils. Furthermore, the main aim in constructing common lesson content was that it would be open-ended enough to give teachers the chance to ask a wide variety of questions related to it in discussion. There are difficulties, however, in trying to standardise lesson materials as was indicated by Dunkin and Biddle (1974, p.443), who added that 'it might be that some process-product relationships do not generalise beyond specific content and context ...'. The materials used in the present study were therefore restricted to cross-cultural topics in social studies.

Nuthall (1970) also has claimed that studies which analyse variations in teacher behaviour without standardising the teaching materials used were less than satisfactory, for the inconsistencies in teaching materials may be responsible for any differences observed. Gall (1970) has indicated that when a researcher tries to determine the effects of a training programme upon teachers' use of questions, constant lesson materials should be used. The studies by Hunkins (1966) and Galassi, Gall, Dunning and Banks (1970) are two examples of the few studies to make use of standardised materials for both pretreatment and posttreatment lessons.

The three sets of standardised material developed for this research, were designed so that they were consistent in the types of objectives set, in the broad social studies concepts required, and in the notes and television content given to teachers: that is, each teacher was given exactly the same materials.

The objectives of the lessons were designed to give teachers a chance to include a variety of question types in class discussion. The objectives given to the teachers are listed in Appendix C (p.264). A wide range of thought processes was covered by the objectives, and these were, in turn, the basis for the achievement test items. For example, the objective '... put forward their [pupils] own ideas

relating to the content studied', was stated to encourage teachers to engage in an open, flexible discussion with their pupils, including the use of divergent questions. Because the lesson materials were about similar topics, the list of objectives used for the three lessons was virtually the same. The order of listing of the objectives was by chance, and no order of priority was therefore implied. Teachers were reminded of this in lesson briefings.

Several broad, abstract concepts were formulated on the basis of the concepts in the *Draft Syllabus in Social Studies: Forms One to Four* (1972), and Shindelus and Durkin (1972), and these were: cultural change, cultural difference, institutions, interdependence, values, and tradition. Appendix C (p.264) lists these more fully.

For each lesson package, a topic was selected which pupils were unlikely to have studied. Preliminary checking earlier in the year in which the data were collected showed that the topics met this requirement. The topics were:

Lesson One: A Tongan village.

Lesson Two: A Hopi village.

Lesson Three: A Temiar village (Malaysia).

When the classes were shown the materials, it was found that a few pupils knew a little about Tonga, but not the village studied, fewer still knew anything about Hopis, and no child had heard of the Temiar people.

Before each lesson — in a briefing session outlined later (Chapter 6) — teachers were provided with printed material which was designed to give them sufficient background knowledge about the village so that they would be able to conduct a satisfactory discussion. That is, the notes were designed to give supplementary information about the village under study so that the teachers could — in a relatively short time — become reasonably knowledgeable about the topic. Appendix C lists the sources of these materials.

The substantive content for each lesson was given to the classes concerned in a specially-designed television presentation. For each lesson, the investigator designed a 14-minute sequence which was recorded in the teaching resources section of the Education Department

of the University of Waikato.

The procedure was as follows. The sequence was explained to the television technician, who advised on the best recording and editing procedures, and suggested modifications. A film was made of a five minute introductory phase, fronted by the investigator, in which basic information about the country and village was presented. For this purpose, use was made of maps and still photographs. A nine minute excerpt of movie film was added to this introduction. Each film was about village life, and was adapted from National Film Library films.

The sequence was then trialled in the pilot school classrooms, and modifications subsequently made. In all cases, the material was re-recorded to improve quality. The time consumption was large, but the advantages for standardisation of lesson materials outweighed this problem. Before the television presentation was finally approved it was subjected to the scrutiny of several Teachers' College social studies lecturers, who confirmed that the material was appropriate for its purpose of promoting open-ended inquiry-type discussion.

CHAPTER 6

DATA COLLECTION AND RESEARCH INSTRUMENTS

OVERVIEW: In this chapter the procedures used to collect classroom data are outlined. One pretreatment and two posttreatment lessons were taperecorded in all sample teachers' classrooms, and on the day following these social studies discussion lessons the pupils were given an attainment test. The instruments used in this experiment are then described. The Rokeach and Zevin scales were administered to measure teacher attitudes; standardised test results were collected as possible covariates and a standardised inference test administered to pupils; and details of how achievement tests were constructed and administered are given. Finally, marking and data classification procedures are presented: a system of classifying teacher questions was developed specially for this study both for use in the in-service training treatment and as a basis for analysing teacher questioning. The categories are operationally defined and the procedures used to train coders to code teacher questions from lesson transcripts are given.

6.1 Procedures for data collection

The experimental design and general procedures for data collection have been described in the previous chapter. Further details of the methods used to gather relevant information in the classroom setting will now be given.

6.1.1 Taperecording classroom lessons

Three observations of teacher questioning were required — one pretreatment and two posttreatment. Before each discussion lesson was taken by the teacher concerned, the investigator took videotape equipment and the lesson materials to each school, at a pre-arranged time, to brief the four teachers; this briefing lasted about 40 to 50 minutes. The procedure for the first lesson, about a Tongan village, was as follows:

1. The notes "Recorded lesson on Tonga' (Appendix C) were

issued, and the teachers asked to read them. These notes outlined the recording procedures to be followed.

2. Lesson objectives and teacher content notes were distributed, and teachers read them (Appendix C).

3. The investigator pointed out to the teachers that the lesson objectives covered a range of potential learnings, and that no order of priority was intended in the way objectives were listed.

4. Teacher queries were answered. In the Tonga briefing, queries most often related to teachers trying to work out what the investigator 'wanted' in terms of a lesson. It was stressed that no preconceived views were held about how the lesson should be taught, and it was also stressed that each teacher should prepare lessons independently of the others taking part in the experiment. Subsequent inquiries showed that this request had been met.

5. Teachers viewed the videotape lesson material, and were encouraged to take notes while they did so, to assist lesson planning. Further teacher queries, usually related to clarifying specific points about the television content, followed.

6. Lesson recording procedures were carefully outlined to the teachers.

7. A timetable for recording the lessons was drawn up. This sometimes involved consultation with the Principal to clarify potential timing clashes. In Intermediate schools, timetables are often complex, but all schools were very cooperative in adapting timetables to ensure sufficient time for the lesson and the same time on the following day for testing.

The procedure outlined above was followed in briefing teachers for the second and third (posttreatment) lessons. Compared with the first briefing, subsequent briefings were shorter, because there were fewer queries about details of recording the lesson. Briefing for each lesson took place about one week before the lesson was taught.

This gave enough preparation time, and meant that all teachers had the same opportunity for planning.

An important consideration in classroom research is the effects that observers might have on the behaviour of the teachers and pupils. Medley and Mitzel (1963) pointed out that observers do influence behaviour, but probably not significantly, as long as everything possible is done to reduce observer effects. In the present study it was assumed that because all teachers knew when their lesson was to be recorded, and all had time to prepare, the lesson would represent a maximised performance level by the teachers, in meeting lesson objectives as they saw fit. After doing extensive classroom tape- and videorecordings, Borg (1972) maintained that pre-arranged recordings probably provided a reliable estimate of a teacher's regular classroom performance. This was verified, Borg claimed, by the fact that teachers had difficulty in emitting teaching behaviours which were not part of their usual teaching repertoire, even when encouraged to do so.

To determine observer effects, Samph (1976) compared observations of ten female teachers when an observer was present with taperecordings from hidden microphones when the teacher did not know the recordings were being made. Using the Flanders Interaction System to code verbal interaction, Samph found that there was a significant difference between teacher behaviour when an observer was present and when there was no observer, on the following variables: Indirect-Direct ratio; teacher use of praise; teacher use of student ideas; and criticism. However, Samph found that the most often used verbal behaviours did not change much from one situation to another. That is, it seemed that the teacher's predominant patterns of verbal interaction were not unduly affected by observer presence.

In the present experiment it was anticipated that observer presence would not unduly affect the verbal interaction that usually took place, and Samph's study lends support for this assumption. To reduce possible effects, the investigator attempted to standardise the meetings held to brief the teachers about the lesson content and recording arrangements; he met the children before classroom tape-recordings were made; and he explained the equipment and procedures to pupils before each lesson.

In all classrooms the lesson recordings followed a standardised procedure which had been developed during pilot lessons conducted in an Intermediate school. Two sets of equipment were needed for lessons and taperecordings: video equipment showing pupils the lesson content, and recording equipment to taperecord the class discussion which followed the viewing. Details of these are given in Appendix D (p.273).

The equipment was set up in each classroom either before school, or during an interval preceding the lesson. The video-cassette player and television monitor were placed on tables at the front of the classroom. Microphones were hung from strings across the room, at a heightof seven feet. Three or four microphones were used, and these were connected to a sound mixer and taperecorder at the back of the classroom. Chairs were arranged in two or three rows in a semi-circle, so that all pupils were in close proximity to the television screen, and close to the teacher in the discussion section of the lesson. Pilot lessons had indicated that furniture and shuffling feet reduced the clarity of taperecordings: to improve the sound recording quality in classrooms without carpet (two-thirds of the rooms), the investigator therefore placed large mats on the floor before the chairs were arranged. Noises such as feet scraping and chair leg scrapes were substantially reduced and, as a result, the sound quality of the tapes was very good, so that transcribing them was easier and more accurate. As the setting-up of a classroom, the discussion lesson and removal of the equipment took at least seventy minutes, the maximum number of lesson recordings made in one day was three.

After a check had been made to ensure all pupils could see the television screen, they were familiarised with the equipment by being shown how it worked, and what was going to happen. This usually took about five minutes. It was noticeable that the children were not distracted by the equipment. There were two major reasons for this: they were already familiar with that kind of technology in their classrooms or homes, and the investigator was already known to the pupils as a result of administering the inference tests. The use of colour television appeared to increase lesson motivation, since about three quarters of the pupils did not have a colour receiver in their home.

In four schools, the recordings were carried out in the teachers' usual classroom. In two schools, however, recordings were made in resource rooms where the classes usually had social studies lessons. All the lessons were therefore held in the pupils' normal social studies location.

All lessons were carried out in the following time-slots: prior to mid-morning interval; after mid-morning interval; or in the first hour after the lunch break. Because of school timetable complexities, it was not possible to randomise times and days of the week for all teachers' lessons. In addition, randomisation would have meant that one investigator would not have been able to collect all lesson data, because of the number of visits to separate schools that would have been necessary. In spite of this, each of the three taperecorded lessons for any one teacher was recorded on a different day of the week, and at a different time of the day. However, the Intermediate school pupils were accustomed to timetable variations, and it was not unusual for social studies to be taken at different times each week. Therefore, it was considered that fluctuations in the timetabling of taperecorded lessons were unlikely to have caused any marked changes in teacher or pupil behaviour. All lessons in each series were recorded within a three-week period, minimizing both history and maturation influences. Teacher illness caused the re-scheduling of one briefing session for teachers at one school, and two lessons. Out of a total of eighteen briefings and seventy recorded lessons, this interference was considered minimal.

6.1.2 Administration of Achievement Tests

The achievement test for each lesson was administered 24 hours after the lesson was taught. The teachers were asked not to remain in the classroom during the test, and all complied with the request. Thus teachers did not know what sort of a test had been given, and would not therefore be able to alter their behaviour in a subsequent lesson in anticipation of particular kinds of test items. Children were told, after each lesson, that a test would be given the next day.

The test, consisting of four sub-sections, took approximately 45 minutes. Each was administered by the investigator and an assistant, who was also a trained teacher. The administration

procedure was standardised during pilot study tests, and practice sessions were held for the two test administrators. All schools had conducted mid-year tests for subject achievement, so the children had had recent experience in test conditions. Children sat in test positions to avoid copying.

The four subtests (see section 6.4 below) were divided into two sections. No time limit was placed on the multiple choice subtests of cognitive memory and higher-order processing, and all pupils were able to complete the 24 items within fifteen minutes. For the evaluative and divergent thinking subtests, there were six statements printed on separate cards. These cards were placed — one at a time — at the front of the class. The test administrator read the card, and the children were given exactly two-and-a-half minutes to make responses to the evaluative cards, and four minutes for each divergent item card. Stop watches were used to ensure exact time for all classes.

Test instructions and a specimen answer sheet are included in Appendix E (p.274).

6.2 Instruments to measure teacher attitudes

Two attitude scales were used to find out whether the in-service treatments had influenced teachers' attitudes. These were the Rokeach Dogmatism scale, Form E; and a scale constructed by Zevin (1969), in a doctoral research programme, to measure teachers' attitudes towards inquiry teaching.

6.2.1 Attitudes — the concept

Allport (in Halloran, 1967, p.14) has defined an attitude as

> 'a mental and neural state of readiness, organised through experience, exerting a directive or dynamic influence upon the individual's response to all objects and situations with which it is related'.

This tendency to react in a certain way, when faced with certain stimuli, means that most attitudes are present, but passive, until aroused by the perception of the object of the attitude. The arousal may result in expression through a variety of behaviours. The

attitudes which human beings hold cover an unlimited field, according to Halloran. Rokeach (1968, p.450) defined an attitude as 'a relatively enduring organisation of beliefs, around an object or situation predisposing one to respond in some preferential manner'.

The significance of these definitions for teacher education is the contention that attitudes are learnt, are modifiable, and therefore subject to change. Also, attitudes vary in degree of intensity. Their importance lies in their predisposing influence on behaviour.

Both attitude measures used in this study are Likert-type scales. The respondents indicate the extent to which they agree or disagree with statements. Oppenheim (1966, p.140) reported that the reliability of Likert-type scales 'tends to be good'. However, a criticism of this method is that it lacks reproducibility; that is, the same score may be obtained in different ways. Total scores can therefore obscure individual differences. However, similar scores by different respondents reveal a similar degree of positive or negative reaction, overall. And in this research, the interest in attitude measurement was in intertreatment group differences, and not so much with how individuals changed.

The two scales were administered to the teachers in the Experimental Skills and Experimental General groups at the start of the experimental treatment (July), and again in December. Control Group teachers also completed the scales in July and again in December.

6.2.2 The Rokeach Dogmatism Scale

This scale was chosen for the present study because Flanders (1970) had seen it as a useful indicator of the likelihood that teachers' classroom behaviour might change as a result of teacher education programmes. In addition it was of interest to see whether teachers who took part in an in-service programme that focussed upon inquiry teaching — and therefore stressed the need for teachers to be flexible, open-minded, and receptive to others' ideas — would change more than teachers who had not been exposed to such a programme.

The Dogmatism Scale was designed by Rokeach (1960, p.19) to 'measure the individual differences in the extent to which belief

systems are open and closed'. The scale was constructed after statements were formulated to reflect various definitions of open and closed belief systems. It was designed so that persons adhering dogmatically to diverse viewpoints should cluster at one end of a continuum, while persons with undogmatic views would cluster at the other end. Form E, which was used in this study, was made up of the forty best items from earlier forms, which gave a scale with increased reliability and improved theoretical relationships (Rokeach, 1960, p.73).

The reliabilities of Form E have ranged from .68 to .93 using the test-retest method (*ibid.*, p.90). Hough (1965) reported a split-half reliability of .86 for Form E, after he had administered it to student teachers. To test the validity of Form E, Rokeach compared high and low scorers, in a series of ideological, conceptual, perceptual, and aesthetic activities. It was found that scores were more attributable to personality than to intellectual ability. Subjects who scored as low dogmatic, were more likely to change their belief systems (Rokeach, 1960, pp.286-9), than subjects who scored as high dogmatic.

Lehmann (1963) used Form E to measure students receptivity to new ideas, their degree of open-mindedness, and their degree of authoritarian attitude. With a relatively large sample Lehmann found that a change to more open-mindedness could be observed, and students in general became more receptive to new ideas from freshman to senior years. Females changed more than males. It was not clear, however, whether these changes would have occurred in altered circumstances, such as not attending College. Also, the reasons for change were not clear.

Flanders (1970) indicated that there was evidence, from studies of the teaching behaviour of Teachers' College students, that students scoring high on the Dogmatism Scale were less likely to make the same changes in teaching patterns as did those with low scores. He considered the Dogmatism Scale as one of the most promising personality scales for research into teacher attitude change in relation to alterations in teaching skills (*ibid.*, p.353).

Johnson (1969) used Form E to find whether the degree of dogmatism of supervising teachers affected the change in dogmatism of student

teaching during teaching practice. The attitudes of fifty-three of the eighty students changed towards those of the supervising teacher. Peck and Tucker (1973) reported that Hough and Amidon had found that student teachers with low dogmatism scores — Rokeach D Scale — changed significantly towards the use of indirect teaching methods after being taught interaction analysis. Equally open-minded student teachers, who had not been taught interaction analysis, and more closed-minded student teachers who had learned interaction analysis, did not change as much.

Student teachers with low dogmatism scores — Form D — have been found to be significantly more proficient in teaching performance than student teachers with high dogmatism scores, or closed belief systems (Zahn, 1967). The finding resulted from a follow-up measurement, after the student teachers had been trained in interaction analysis. Kirk (1967) also found that low scores — Form E — were positively related to teaching practice ability in student teachers, as measured by tutor ratings of teaching practice performance.

The Rokeach Scale thus seemed to be potentially useful, in the present study, in identifying teachers most likely to change their questioning behaviour.

6.2.3 The Zevin Attitude-to-Inquiry Questionnaire

The Zevin scale was chosen for use in this study because it seemed to offer interesting possibilities for measuring teachers' attitudes towards a number of important components of an 'inquiry-based' classroom, and for finding out whether or not an in-service programme would change those attitudes. Also, the scale was constructed specifically for use in the field of social studies classroom interaction, and although it has not yet been widely used in research, Zevin himself, and Hurst (1974), have used it to measure the effects of in-service courses focussing upon inquiry teaching on teacher attitudes towards inquiry.

As just stated Zevin (1969) constructed the instrument to measure teacher attitudes towards inquiry teaching. The questionnaire consists of twenty-five items (the full scale is shown in Appendix F, p.306) in the form of statements, which were adapted from the Minnesota Teacher Attitude Inventory. Teachers respond to each statement by means of a

five-point Likert-type scale, ranging from strongly agree, to neutral, to strongly disagree.

Zevin maintained (1973, p.313) that an orientation towards inquiry teaching is demonstrated by teachers who agree with statements favouring 'openness, flexibility, freedom of expression, and the stimulation of thought in the classroom'. Agreement with statements relating to rigidity and authoritarianism is 'assumed to indicate non-inquiry orientation'. The highest possible score is 125 and it was also assumed that the nearer a score was to 125, the stronger would be the teacher's orientation towards the use of inquiry methods. It must be noted that the reported test-retest reliability of .76 (Zevin, 1973, p.313) for individual scores is rather low.

In a study to measure the effects of in-service courses on teachers' classroom behaviour, Zevin (1969) found that teachers who studied inquiry methods in social studies increased their mean score by 7.3 points from pretreatment to posttreatment, while a second group of teachers who had attended an in-service course not focussing upon inquiry methods, advanced their pre- to posttest mean by only 1.3 points. However, the mean of the first group was initially 7.5 points higher than the second, at the start of the in-service programme. The research provides some evidence, nevertheless, that a course which uses inquiry modelling by instructors can influence attitudes-towards-inquiry in a positive direction.

The only other study located that used the Zevin scale was one by Hurst (1974), in which two groups of twenty student teachers were taken through a five-week course to train them in the use of inquiry techniques suitable for elementary school classrooms. The course was not directed at social studies in particular, but the Zevin scale was administered before and after the course, and Hurst found that the two groups — each of which had a different kind of course on inquiry methods — increased their scores by an average of four points. The scores of a control group of the same size did not alter significantly. Hurst noted that, in all three groups, the student teachers already had positive attitudes towards inquiry, as measured by the Zevin scale, before the inquiry course.

6.3 Standardised pupil tests

Three standardised tests were used in this study: The *Progressive Achievement Test (PAT): Reading Comprehension*; the *Otis Intermediate Intelligence Test*; and the *Social Studies Inference Test*. The first two tests had already been administered by the schools, as part of their routine testing procedures. The results were used as covariates in an analysis of covariance procedure which will be outlined in Chapter 7. The inference test, which was administered by the investigator, was used to find out whether or not pupils' ability to make inferences had been affected by changes in teachers' questioning behaviour.

6.3.1 *PAT: Reading Comprehension*

This forms part of a series of standardised tests developed for use in New Zealand schools, and now widely administered. The Reading Comprehension test was designed to measure both factual and inferential comprehension of prose material (Elley and Reid, 1969, p.5). Reliability coefficients obtained by correlating equivalent forms, ranged from .83 to .89. Validity estimates for Form One pupils were made through correlations between Comprehension and: Otis Intermediate Intelligence A (.79); PAT Vocabulary A (.79); ACER Intermediate Intelligence D (.69); and the Academic Promise Test (.68). Standardised 'level' scores and percentile ranks were developed from raw scores.

The test was considered suitable as a possible covariate for measuring posttreatment differences in achievement test results in cognitive memory and higher-order thinking subtests, because of the expected high correlation between these three measures.

6.3.2 *Otis Intermediate Intelligence Test*

The Intermediate level series of the Otis Tests of Mental Ability is widely used in New Zealand to test children prior to entry into Intermediate schools; that is, in their Standard Four year. The tests were aimed at providing an indication of 'a pupil's acquired facility in a broad variety of reasoning skills ...' (Test Manual, 1969, p.1). New Zealand norms were developed from a 1968 representative sample of 5630 children between the ages of 9 to 15 years. Most reliability

studies of the test have produced coefficients close to 0.9. The standard deviations range from 13 to 15 points, and the standard error of measurement is four to five raw score points.

Form A had been administered by the Intermediate school personnel to all pupils used in this study, in the pupils' Standard Four year. The scores were collected by the investigator, to be used as a possible covariate for the statistical design outlined in Chapter 7.

6.3.3 *Social Studies Inference Test*

The test is a New Zealand version of the Inference Test developed in the Taba Project (Taba, 1966; McNaughton, 1974). The instrument was designed to test inferential thinking, and 'gave information on children's tendencies either to mark over-generalizations as appropriate responses or to be over-cautious about accepting valid conclusions' (McNaughton, 1974, p.24).

The Inference Test (see Appendix G, p.309) consists of eleven stories, followed by a number of statements. Pupils respond to the statement as 'probably true', 'probably not true', or 'can't tell'. Stories were considered relevant to New Zealand social studies by the test constructors. Percentile ranks were calculated from the raw scores of over one thousand pupils at the Standard Two, Standard Four, and Form Two levels. Two dimensions were standardised: inference, and caution. Kuder Richardson Formula 21 reliability for Form Two Inference was .76, with .83 for caution scores. Low correlations with Otis Intermediate A (.30 to .40), indicated that the test measured abilities which varied from those measured by the Intelligence Test (McNaughton, 1974).

The administration of the Inference Test followed the procedures outlined in the manual (McNaughton, 1974), and the average test time was forty-five minutes.

6.4 Constructed Achievement Tests

Three achievement tests, based on the television lesson materials, were constructed; one for each pretreatment and posttreatment lesson. Each test was divided into four subtests which were designed to measure

different dimensions of social studies achievement: cognitive memory; higher-order (convergent thinking); evaluative responses; and divergent responses. These dimensions are consistent with the Guilford (1967) model of operations of thinking, and consequently with the question classification system developed in the present study to measure teachers' questioning behaviour. The sub-test results were to be analysed separately, to find whether changing patterns of teacher question-asking affected pupil achievement on a range of thinking operations.

Dunkin (1976) has recently stressed the importance of these considerations in testing in process-product research. He stated that if product measures (test results) are removed from the content of lessons, there is less chance of identifying process-product relationships. Also, the wider the range of cognitive levels that are measured, the greater will be the opportunity of getting better evidence of relationships between process variables and a variety of outcomes.

The construction and use of the achievement tests represented an exploratory aspect of the present research, because very few studies have attempted to relate changes in teacher questioning to pupils' test achievement, and no research was located in which the effects of teacher questioning were investigated by means of subtests specifically designed to measure various pupil thinking abilities.

A pilot test of about sixty items was constructed, and administered on the day following the pilot lesson, which was to be the period of lesson-test delay in the main data collection phase. The pilot school was located in a town near Hamilton City, and the school was not used in the research sample. Two classes were used, and fifty-four pupils completed all three pilot tests (26 in Form One, and 28 in Form Two). The classes were heterogeneous in measured intelligence.

6.4.1 Cognitive memory and Higher-order subtests

The items for these subtests were multiple choice. Lindvall and Nitko (1975) reported that tests using multiple choice items 'are usually more reliable than those containing other types', and 'can be used to evaluate a greater variety of abilities than can other types'

(p.58).

The procedure for constructing the multiple-choice items was as follows. The items intended to measure cognitive memory were written so that a pupil would be able to answer correctly simply by remembering information which he was given in the television presentation. Therefore all pupils had the same informational base, regardless of any discussion which they undertook following their viewing of the material.

In the Higher-order subtest the pilot items were designed to get pupils to use thinking which went beyond the knowledge level (Bloom, 1956). The pupil would not be able to answer this type of question solely by remembering a single piece of information. Rather he would have to re-combine factual information by using thinking involving one or more of comprehension, application, analysis, and synthesis abilities.

When the pilot items had been written they were shown to three lecturers who were familiar with both Bloom's taxonomy and the social studies field. They were also shown the television presentation and then asked to say whether each test item measured cognitive memory or a higher level of thinking. An item was retained in the pilot test only when perfect agreement was reached.

Item analysis was computed from the highest 27 per cent and lowest 27 per cent of pilot study raw scores, after the technique recommended by Gronlund (1968). Item difficulty indices and item discrimination indices for the items finally selected are shown in Table 6.1 (p.187). The difficulty indices for multiple-choice items ranged from .37 to .83 (Tonga), .29 to .82 (Hopi), and .37 to .83 (Temiar). The mean difficulty indices were .60 (Tonga), .60 (Hopi), and .60 (Temiar).

Discrimination indices ranged from .27 to .73 (Tonga), .27 to .93 (Hopi), and .27 to .73 (Temiar). The mean discrimination indices were .42 (Tonga), .47 (Hopi), and .41 (Temiar). These difficulty and discrimination ranges fall within the limits recommended by Gronlund (1968), and when these figures are related to the means and standard deviations for the final form of each test (Tables 6.2 and 6.3), a high degree of consistency between the parallel tests is indicated.

It should be noted, however, that the statistical analyses of test scores, to be undertaken later in accordance with the experimental design, were to find whether there were differences between ES, EG and C pupils for each test, and not to compare directly the results of one test against the others.

To check on the validity of the Cognitive memory and Higher-order thinking subtests combined, correlation coefficients were calculated between the results of the three parallel tests, and between those of each test and the Otis Intermediate A intelligence quotients. The results are shown in Table 6.4, and indicate a range from .68 to .78. These levels were considered satisfactory for achievement tests of this kind. Content validity was checked by verification from three lecturers in social studies that items did indeed relate to the substantive lesson content as was indicated above.

The Kuder Richardson Formula 21 was used to calculate reliability. Results were: Tonga test .80; Hopi test .82; and Temiar .82. Tuckman (1975) indicated that this level of reliability was highly satisfactory for non-standardised tests.

A split-half technique was also used to check reliability. With a Spearman-Brown formula correction (Tuckman, 1975, p.261), the 24 items finally selected for the Tonga test had an internal consistency coefficient of .82. Because of the similarity of item types and item analyses results, the Hopi and Temiar tests could be assumed to reach a similar level.

A table of specifications for the 24 items in each test is shown in Table 6.5. The achievement tests themselves are reproduced in Appendix E.

6.4.2 Evaluative and Divergent thinking subtests

The development of these subtests represented another exploratory aspect of this study. The free response items used required pupils to write as many responses as possible in a prescribed time.

The desirability of pupils giving their own ideas, opinions and reasons has often been referred to in the literature on social studies education (Beyer, 1971; Ploghoft and Shuster, 1971; Lawton and Dufour, 1973). However, testing in social studies has tended to concentrate upon convergent thinking and memory skills (Klein, 1972). The need

for greater stress on creativity and divergent thinking in classrooms generally has been well established (Getzels and Jackson, 1962; Torrance, 1962; Wallach and Kogan, 1965), but according to Flanders (1970), most classrooms concentrate on convergent thinking rather than divergent thinking. No research studies could be located which had used free response evaluative or divergent thinking test items to measure social studies achievement related to teacher questioning behaviour.

There is controversy in the literature as to the extent to which creativity or divergent thinking test scores are positively related to achievement test scores (Getzels and Dillon, 1973). It was claimed in one study that children who do well on divergent thinking tests contribute more in class discussion and write more imaginative stories than other children (Vernon, 1973). It was hoped that the present study would provide some evidence as to whether or not children who are encouraged to think divergently also write down more ideas in free response test items.

The evaluative thinking items were designed to encourage pupils to give reasons for agreeing or disagreeing with a value statement. To answer, a child was required to use 'evaluative thinking' as this is defined by Bloom (1956) and Guilford (1959). Reasons were scored if they were given on the basis of evidence from the lesson content. Marking procedures are discussed below in Section 6.5.

Divergent items were based on the 'Just Suppose' section in the *Torrance Tests of Creative Thinking* (Torrance, 1966). The 'Just Suppose' items were designed to elicit a higher degree of spontaneity and to be more effective with children' (*ibid.*, p.13). The pupil was confronted with an unusual situation which required him to utilize his curiosity and to 'play with' various possibilities and imagine all the things that would happen as a consequence (*ibid.*, p.13).

In the present study, pupils were presented with a situation which did not exist, and asked to write down all the ideas they could about how the lives of the villagers in the previously-seen television presentation might be affected if they were placed in such a situation. For example, they were asked 'Just suppose oil was discovered in the village'. Scoring was based on ideational fluency and flexibility. These measures were used by Torrance (1962) to score divergent

thinking qualitatively as well as quantitatively. Ideational fluency requires a child to give plausible ideas, while flexibility measures the extent to which a child can 'shift' his thinking to produce a number of ideas about the same stimulus. It was assumed that, in order to make responses, pupils would need to recombine knowledge they had about the cultural setting in ways which for them were new and original. A table of specifications for the evaluative and divergent thinking items is shown in Table 6.6. Each test contained three evaluative thinking and three divergent thinking test items.

Validity is difficult to establish for test items that claim to measure divergent abilities, since there are an infinite number of ways a person can behave creatively (Torrance, 1962). For free response evaluative thinking items, validity is equally hard to establish. As will be indicated in more detail shortly, factor analyses of the pilot study results produced evidence that the evaluative and divergent thinking items were probably measuring dimensions other than those measured by the cognitive memory and higher-order thinking multiple-choice items. Also content validity was checked by consultation with five Teachers' College social studies lecturers. These judges were asked to say which statements reflected evaluative thinking, and which reflected divergent thinking in relation to the materials used for lessons. Perfect agreement resulted.

6.4.3 Thinking operations incorporated in the subtests

One of the intentions of the subtests was that they would measure a range of thinking abilities. However, the number of test items in each subtest was comparatively small, as was the number of pilot pupils (N = 54). Therefore, to determine whether or not the subtests were measuring different dimensions of thinking operations within the domain of social studies, the raw scores of the pilot pupils were subjected to a principal components factor analysis. A computer programme from the *Statistical Packages for the Social Sciences* (Nie *et al.*, 1975) was used. The pilot test items for each subtest were divided into split-half groups to produce a more powerful check on factor loadings, for if loadings on the two halves of a subtest were similar, then it could be assumed that both halves were probably

measuring the same thinking dimension. The subtests in the pilot study were: Cognitive memory (20 items); Middle-order thinking (comprehension) (14 items); Higher-order thinking (10 items); Evaluative thinking (4 items); and Divergent thinking (4 items).

Table 6.7 shows the results from a varimax rotated factor matrix for the pilot subtests' split halves. From the results, four factors were found with eigenvalues of one or greater, and these four factors accounted for 68.1 per cent of the variance. The table shows that Factor One had a high loading on both halves of the Evaluative (.65 and .81) and Divergent thinking subtests (.78 and .66). This seemed to indicate that to answer items in these two subtests pupils were using some thinking dimension which was different from the abilities they used in answering the other subtest items.

Factor Two loaded highly on both halves of the Cognitive memory subtest (.80 and .83), and also loaded moderately on the Middle-order subtest halves (.29 and .36), thus indicating the probability that memory ability was the main contributing dimension in the answering of these items. However, Factor Two also loaded on Divergent thinking, B half (.41). Factor Three loaded low on all variables except Evaluative thinking, B, (.36), and Divergent thinking, B, (.30).

By loading on Middle-order thinking (.32 and .79) and Higher-order thinking (.86 and .30), Factor Four gave an indication that children were using a different set of abilities in these subtests, than they were in the Cognitive memory, Evaluative thinking and Divergent thinking subtests.

On the basis of this evidence, and taking into account the item analysis data, a final form of the Tonga test was constructed. This in turn was factor-analysed, with results as shown in Table 6.8. Five factors were identified: the first two accounted for 71.9 per cent of the total variance, and three others accounted for a further 28.1 per cent. The loadings of Factors Three, Four and Five, must however, be viewed as suggestive only as their eigenvalues fell below 1.00. The significance of the loadings is that they indicated that each of the subtests was indeed measuring a different dimension of thinking. Factor One loaded strongly on the Evaluative thinking subtest (.95); Factor Two loaded highly on the Cognitive memory subtest (.95); and Factor Three loaded at .95 on the Divergent thinking subtest.

While the number of pupils available was small, the results of these factor analyses suggest that there was justification in relating the subtests to the four primary thinking operations of the Guilford (1956) Structure-of-Intellect Model, with Cognitive memory representing a combination of cognition and memory in Guilford's terms.

It will have been noted that, in the pilot study, items above the cognitive memory level had been subdivided into thinking categories based, respectively, on the comprehension and application, analysis and synthesis categories of Bloom's Taxonomy. However, it proved difficult to establish content validity for these subtest items because raters (Teachers' College social studies and evaluation educators) could not achieve a sufficiently high degree of agreement in assigning items to one or other of the two divisions. In view of this lack of agreement, it was decided for the final test form to combine the Middle and Higher-order thinking subtests into one.

6.5 Achievement Test marking

The subtests of multiple-choice items were marked by the investigator using a marking key. Reliability of marking was established by getting the assistant test administrator to mark randomly selected answer sheets. The agreement was perfect.

The Evaluative thinking and Divergent thinking subtests were marked by the investigator in the following way. For the Evaluative thinking items, one point was given for every reason that had supporting evidence obtainable either from the film materials or the ensuing class discussion. The evidence from the discussion was checked by reference to the transcript of the lesson. Each pupil was given a summed score from the three Evaluative thinking items.

The Divergent thinking items were marked for ideational fluency and flexibility (Torrance, 1962). To score a point an answer had to have two qualitative requirements. First, the answer had to have some degree of plausibility. This meant that if, in response to a question like 'How might the lives of the Temiar people change if oil was discovered there?', a child put 'Get the Martians to guard it', no

point would be scored. However, the answer 'Try to sell the oil for hunting guns' has plausibility and would score a point.

Second, only separate and different ideas were scored. For example if, as answers to the above question about oil, a child wrote 'The oil might get all over their crops', and 'The oil might get all over their traps', only one point would be given because there is really only one basic idea involved — oil getting all over something. Pupils again received a summed score from the three Divergent thinking items.

To establish reliability of marking for the free response items, the marker's scoring was compared with the marking of randomly selected answer papers by three lecturers in social studies who acted as raters. A short — 80-minute — training session was held. The raters were shown the lesson materials for one package, and the aims and techniques of scoring items were explained. Using practice papers, responses were discussed until all raters agreed about marks allocation. A formal agreement check was then carried out, with each of the three raters and the investigator independently marking four randomly chosen papers. A high degree of inter-rater agreement was achieved. In the case of nineteen out of the twenty-four items checked, all four markers gave identical scores. Interpreted another way, the total marks for the four test papers for Evaluative thinking items were: raters three and four, 42 points; rater one, 44 points, and rater two, 41 points. Totals for Divergent thinking items were: raters one and four, 54 points; and raters two and three, 55 points. These figures indicate a high level of agreement, and it was assumed that the marker (investigator) could achieve a high level of consistency over the marking of the total pupil sample. A subsequent agreement check with four test papers, using the same raters, showed that the high agreement levels were in fact maintained, with less than 10% difference between the highest and lowest rater scores for Evaluative thinking and Divergent thinking.

For the marking of the Social Studies Inference Test, the procedures outlined in the test manual were followed. Scores were converted to percentile rankings for Inference and Caution (McNaughton, 1974). No marking error was found when twenty randomly chosen answer sheets were marked by the assistant test administrator, and compared

with the marker's papers.

6.6 A system of classifying teacher questions

6.6.1 Introduction

A specially constructed classification system for teacher questions asked in social studies lessons was devised for this study. The system focussed upon cognitive abilities included in Guilford's Structure-of-Intellect model (Guilford, 1956). The system was required to be suitable for use in the in-service skills training programme, and also as an observational instrument.

The major underlying assumption in developing the system was that, in order for teachers' questioning patterns to be influenced by in-service work, the question types needed to be stated in straightforward, operationally-defined, terms. There is insufficient time, in short in-service courses, to learn and practice complex systems. Furthermore, it was considered preferable to start out with a relatively simple system, to find out whether teacher behaviour would change as a result of the in-service experience. If change did occur, then more complex questioning behaviour might be the subject of further research.

6.6.2 Theoretical base

Guilford (1956) hypothesised a three-dimensional structure for intellect: operations, contents, and products. Guilford claimed that intellectual functioning involved performing a particular type of cognitive operation upon a particular type of content, to produce a particular type of product. The five types of operations postulated were cognition, memory, convergent production, divergent production, and evaluation. One hundred and twenty intellectual abilities were suggested in the theory, and Guilford (1970) claimed that over eighty had been identified through factor analyses. Furthermore, each cell in the model might contain more than one ability.

The present research focussed upon the operations section of the Guilford model, which relatively few classroom studies have used as the theoretical basis for research instruments (Dunkin and Biddle, 1974).

1. Cognition, the first thinking operation, has been defined to include the processes involved in 'the discovery of information and the rediscovery or recognition of information' (Guilford, 1959, p.359).

2. The memory operation involves the retention of information which the learner shows by reproducing or recalling previously acquired knowledge. This operation has essentially the same meaning as the Bloom (1956) category of 'knowledge'.

3. Convergent production 'proceeds toward one right answer, that is to say, a determined or conventional answer' (Guilford, 1959, p.359).

4. Divergent production, on the other hand, is 'the kind that goes off in different directions. It makes possible changes in direction in problem solving and also leads to a diversity of answers, where more than one answer may be acceptable' (Guilford, 1959, p.381). Creative thinking is an aspect of divergent thinking.

5. Evaluative operations 'have to do with testing information and conclusions as to their suitability, acceptability, goodness, or correctness. This involves the question of standards or criteria' (Guilford, 1959, p.390). The category is similar to the evaluation category of Bloom (1956).

The Aschner-Gallagher Classification System was the first major attempt to utilize the Guilford cognitive operations for classroom research (Aschner *et al.*, 1965). The original Aschner-Gallagher system was designed to classify thought processes which occurred in classroom interaction. The system was comprehensive, in that it attempted to cover all teacher and pupil talk. Questions were part of the system. All talk was classified within each of four primary thinking types: cognitive memory, convergent, divergent, and evaluative. A fifth category — routine — covered managerial-type talk. Guilford's cognition and memory categories were linked into one category, because of 'the difficulty in differentiating them in class interaction and because both represent, from our point of view, unproductive thinking operations' (Gallagher, Aschner, and Jenné, 1967, p.19).

In analysing the results of a study of gifted children, the same researchers made a distinction between teacher questions and teacher statements. They claimed that the questions teachers asked were indicative of the type of thought processes asked for on the part of pupils.

6.6.3 The question categories

In the present study, four basic categores of the Aschner-Gallagher system have been included. Some re-definition was made within the evaluative and divergent thinking categories to clarify their role in social studies lessons. Open and closed sub-categories were built into three of the categories, and two additional ones were developed: extension and grounding questions. The system was thus designed to record the cognitive aspects of teacher questioning.

The four primary categories used were:

A Cognitive memory questions.
B Convergent thinking questions.
C Evaluative thinking questions.
D Divergent thinking questions.

The three secondary categories were:

E Grounding questions.
F Extension questions.
G Routine questions.

The first three primary categories were then divided into two sub-categories: 'open' and 'closed'.

The system can be expressed diagrammatically:

Classification System for Teacher Questions

Primary Categories				Secondary Categories		
Cognitive memory	Convergent	Evaluative	Divergent	Grounding	Extension	Routine
CLOSED	OPEN					

The sub-categories closed and open were derived from the categories 'narrow' and 'broad' used by Amidon and Hunter (1967).

Closed questions ask for one acceptable answer, whereas open questions invite two or more acceptable answers. The inclusion of the open and closed sub-categories was considered important for social studies inquiry learning because, in conducting the experimental in-service courses, it was intended to point out that by using large numbers of open-ended questions, teachers can give pupils more opportunities to use thinking which goes beyond simple recall. The need for teachers to make greater use of open questions had been shown by Nuthall and Lawrence (1965), who found that questions requiring simple 'yes' or 'no' answers were asked frequently by teachers. In this study the two sub-categories were therefore included for teaching purposes rather than as part of the statistical analyses of teacher questioning.

Pilot lessons for this study revealed that most Cognitive memory questions were closed, Convergent questions were made up of approximately equal numbers of closed and open categories, Evaluative thinking questions were mostly open, and Divergent thinking questions were necessarily all open.

Operational definitions of the categories were formulated as follows:

A. Cognitive memory questions. These questions are limited to the lowest level of thinking. They seek the recall or recognition of facts and definitions in this study, usually something the pupils have been told in the television material, or by the teacher.

A.1 Closed. Calls for one correct answer.
Example: 'How many islands are there in Tonga?'

A.2 Open. Calls for two or more answers.
Example: 'What kinds of vegetables do the Tongan villagers grow?'

B. Convergent thinking questions. This kind of teacher question 'represents the analysis and integration of given or remembered data' (Gallagher *et al.*, 1967, p.19). The question is directed toward one or more expected answers, because it is framed within a tightly-structured knowledge system, and an individual's response must be channelled in that direction.

Convergent thinking questions are cognitively more demanding than Cognitive memory questions, because they require the

answerer to put facts together to construct an answer. Thinking operations involved include explaining, stating relationships, associating and relating, and comparing and contrasting.

B.1 Closed: Calls for one correct answer.

Example: 'What characteristic did the vegetables grown by the villagers have in common?'

B.2 Open. Calls for two or more correct answers.

Example: 'For what reasons did the Tongans wear light clothing?'

C. Evaluative thinking questions. In the Aschner-Gallagher system, the Evaluative question represents the highest level of thinking. It 'deals with matters of judgment, value, and choice and is characterised by its judgmental quality' (Gallagher *et al.*, 1967, p.20). The answerer is required to organise his knowledge, formulate an opinion, and take a self-selected position.

For the system employed in the present research, a more flexible definition was devised, because it was designed for use with teachers of eleven to thirteen year olds. The Evaluative thinking question gets children to use skills of making value judgments, judging opinions, justifying a value choice, and giving opinions. For closed questions, reasons do not necessarily have to be given.

In pilot lessons, some difficulty was found in classifying several questions as either Cognitive-memory or Evaluative thinking because in those cases it was hard to tell whether the pupil who answered the question had been given sufficient evidence upon which to form an opinion. In situations where criteria had been given in the lesson content, the question might require little more than a memory response.

C.1 Closed. These questions call for a 'yes-no' or 'agree-disagree' answer. However, the response is regarded as being more than memory, because the pupil has to make a mental decision about an answer, from previous information. Reasons need not be given. This question often leads to an Extension or Grounding question.

Example: 'Do you agree, John, that the people work hard?'

C.2 Open. The questions seek value judgments with reasons, and require more than simple 'yes-no' type answers.

Example: 'What reasons do you have for liking the villagers' way of life?'

D. Divergent thinking questions. Divergent thinking questions call for the use of originality and imagination by the pupils. The questions are thought-provoking, and encourage the responder to organise known data into new patterns that he (and perhaps the class and teacher) was not aware of previously. However, every response does not have to be original in that way. The Divergent thinking question is broad and open, and calls upon a range of possible answers. The teacher has no necessarily fixed answers in mind when he asks the question.

Prediction, inference, and hypothesising are important thinking processes involved in answering Divergent thinking questions. Bloom's (1956) synthesis category is involved:'this is the category in the cognitive domain which most clearly provides for creative behaviour on the part of the learner' (*ibid.*, p.162).

The following definition summarises the above qualities:

> Such a question has no 'right' answer. It is an open-ended question requiring students to use both concrete and abstract thinking to determine for themselves an appropriate response. Students are free to explore the problem in whatever direction they prefer; they are asked to think creatively, to leave the comfortable confines of the known and reach out into the unknown.
>
> (Allen *et al.*, 1969, p.53)

Example: 'How might the villagers' lives change if oil was discovered there?'

E. Grounding questions. Massialas and Sprague (1974) indicated the importance of grounding in social studies lessons, 'to probe, test and compare their [pupil] positions and hypotheses' (p.17). The operations of grounding include 'using evidence to validate hypotheses or positions' ... 'exploring logical consequences of positions', and 'generalising' (p.17). Allen *et al.*, (1969, p.20) and others have called this kind of question 'probing'.

Grounding questions usually occur when a teacher calls upon a pupil to follow up an answer to a primary question. The question is frequently a 'why' question.

Example: 'Why do you say that a plough is essential?'

F. Extension questions. The operational definition of this type is based on the classification formulated by Taba (1966): 'questions which seek further ideas or clarification of the preceding idea. It may also be used to designate continuity of the same idea across several thought levels' (*ibid.*, p.138). The question usually follows an answer to a primary question, and occurs when a teacher wants to encourage children to think beyond the simplistic level.

Example: 'Can you expand your answer more?'

G. Routine questions. These questions 'include procedural matters such as management of the classroom, structuring of class discussion, and approval and disapproval of an idea or person' (Gallagher *et al.*, 1967, p.21). In using these questions, the teacher keeps 'control' of the class discussion to ensure that participation by pupils is more effective. Excessive use could, however, hinder participation.

Examples: 'Would you say that again please?'
'Does anyone know the answer?'
'Why were you talking, John?'

The questions included in the classification system just outlined were considered to reflect a range of types, appropriate for inquiry lessons in social studies. The question types were assumed to foster a range of thinking operations in pupils of Intermediate school age. It was also assumed that teachers would be able to learn how to frame questions of each type through the SPRITE in-service course outlined in Chapter 5, and that the skills learnt would transfer to classroom practice.

6.7 Coding teacher questions from lesson transcripts

To establish whether or not any changes occurred in teaching questioning behaviour, it was necessary to code each teacher question in the lesson transcripts into one of the categories in the system designed for this study.

The investigator transcribed the first twenty minutes of the pupil-teacher talk which had been taperecorded in each of the twenty

to twenty-five minute lessons. All teachers taught for at least twenty minutes. The accuracy of transcription was checked by a colleague, who listened to six 5-minute samples from the tapes, and read the transcripts of those samples. No error was found that would have altered the substantive meaning of any teacher question, which was the main unit of analysis. The clear quality of taperecordings obtained from the standardised recording procedures referred to previously was the main contributing factor to this consistency. The coding was done by two independent coders, who were trained by the investigator, in a seven-and-a-half hour training programme. The coders were university graduate students who had completed, in the same year, a Masters course in the 'Analysis of Teaching'. That course had involved the study of several classroom interaction systems, and practice in coding some of them. Basic interaction terminology and procedures were therefore familiar to the coders. One of these graduates had had coding experience for another researcher. Both graduates were highly recommended for this coding task by the teacher of the course and were able to devote full time to the coding.

The total training time of seven-and-a-half hours was based upon five-and-a-half hours of intensive training, and two hours of individual practice. Research studies reviewed indicate a wide variation in the training time necessary for coders to reach satisfactory levels of agreement in coding various categories. The time required depends upon the complexity of the categories and upon the level of inference involved. Young (1969) organised a twenty-hour training session for eight coders to code videotapes, using the Flanders Interaction Analysis Categories. However, much of the time was spent viewing videotapes, rather than in intensive training. Inter-coder reliability coefficients of agreement ranged from .77 to .95. Also using the Flanders' system, Ryan (1969) reported inter-coder reliabilities for two graduate coders trained by himself in three sessions: .67 after the first session, .86 after the second, and .97 after the third session. Baker (1970) trained coders for several hours to use the Flanders' system and achieved reliabilities somewhat lower than Ryan's and Young's: .75. However, whereas Ryan used tapescripts, Baker used a computer system in which coders listened to a tape of interaction, and pressed impulse buttons to code. Higher initial coder error might be expected from

such a procedure. Both Ryan and Baker used the Smith *et al.* (1964, p.41) formula to compute agreement coefficients, and Young used the Scott formula (Scott, 1955).

The teacher question system used in this study was considered to involve — when coding — a higher level of inference than the Flanders' system used in the above studies. Therefore, an agreement level of .75 as measured by both the Smith and Scott formulas, was aimed at.

During the training programme, which is described in Appendix H (p.325) two reliability checks on inter-coder agreement were carried out. The two coders coded the same transcript independently, and the Smith *et al.* formula was used to compute reliability coefficients. The sub-categories 'open' and 'closed' were removed for this purpose, because there was some difficulty for coders in deciding whether a question asked for a single answer, or more than one answer; and also because these sub-categories were mainly included in the classification system for teaching purposes in the in-service programme rather than as variables for statistical analysis. The coefficients for the second check — after five hours training — were: .73 for cognitive memory; .75 for convergent; .80 for evaluative; .86 for grounding; .89 for extension; and .86 for routine. There was no divergent question.

The investigator believed that little improvement could be expected by further intensive training. Therefore coding of the research tapescripts proceeded. The coders had no knowledge of the research design, nor of the teachers' identities.

During the coding of the transcripts, four transcripts were used as reliability checks for inter-coder agreement. The transcripts were randomly selected, and placed in the folders of transcripts given to coders. The reliability checks represented every ninth tapescript. Table 6.9 shows the results of these four checks, using the Smith *et al.* (1964) formula. The coefficients range from .68 to 1.00, for individual question categories, and from .78 to .89 for mean agreements over all categories.

The Scott Index was also used to compute inter-coder reliability coefficients (π) from the same tapescripts. The formula is:

$$\pi = \frac{Po - Pe}{1 - Pe},$$

where Po (observed percentage agreement) represents the percentage of judgments on which the two analysts agree when coding the same data independently; and Pe is the percent agreement to be expected on the basis of chance.

$$Pe = \sum_{i=1}^{k} pi^2,$$

where k is the total number of categories and pi is the proportion of the entire sample which falls in the i^{th} category (Scott, 1955, pp.323-4).

The Index can be interpreted as the extent to which the coding reliability exceeds chance (Scott, 1955). By using squared proportions over all categories, the Index overcomes a disadvantage of the percentage agreement method; that is, where there are, for example, just two questions of one category, a single disagreement produces a level of agreement of only .50. Table 6.10 shows that the inter-coder reliability coefficients ranged from .67 to .84, using the Scott Index.

Three weeks after the transcripts had been coded, a further check was made, to measure intra-coder reliability; that is, whether each coder consistently coded questions into the same category, using the same transcript. The Scott Index coefficients were .71 and .84 for the two coders. The Smith *et al.* (1964) method coefficients are given in Table 6.11 and show that one coder's mean coefficient was .77, and the other's .86. The main factor contributing to the lower coefficient, was that several questions had been coded as 'Extension' in the first transcript, but as 'Convergent' in the second, and two questions classified as 'Evaluative' in the first transcript had been coded as 'Divergent' in the second. Overall, both inter- and intra-coder reliability levels reached were considered to be satisfactory for a question classification system requiring a moderate degree of inference.

In Appendix II a full outline of the training programme for coders is presented, and a manual of coding instructions is given.

6.8 Summary

Two aspects of this experiment reported above were exploratory in the sense that very little research has previously been carried out in these fields. One was the construction of three achievement tests, each comprising four subtests which were designed to measure different thinking abilities related to the Guilford (1956) Structure-of-Intellect Model. The subtests for Evaluative thinking and Divergent thinking were based on free responses to statements about the lives of the people in the cultures being studied, and this free response technique has been used very little in classroom research.

The other aspect was the construction of a system for classifying teacher questions in social studies. The system represented an attempt to include categories which could, firstly, be easily learnt by teachers in a relatively short in-service programme and, secondly, would provide the potential for the asking of a range of questions which would promote pupils' inquiry learning in the classroom.

CHAPTER 7

PRESENTATION OF THE RESULTS

OVERVIEW: The research hypotheses outlined in Chapter 4 are stated in precise form in this chapter, and the results of the statistical analyses for each hypothesis are presented. For teacher variables of questioning, verbal interaction patterns, amount of talk, inquiry scores and teacher attitudes, multiple analysis of variance and one-way analysis of variance were used to analyse the data. For pupil variables — test scores — analysis of covariance and one-way analysis of variance were used. Significant differences between teacher groups, and between treatment classes were found for several variables.

It was stated in Chapter 5 that the major aim of this study was to find out whether a skills-based in-service course would lead to greater changes in certain teacher behaviours than would result from a general course, and whether such changes would be associated with corresponding changes in pupils' attainment.

A number of relevant variables were identified earlier, and these can be grouped into two clusters: those relating primarily to teacher behaviour and those relating mainly to pupil behaviour. Also, six major groups of hypotheses were formulated; these will now be re-specified (in null form) and the results for each presented.

7.1 Immediate post-treatment effects

This group of hypotheses may be stated as follows:

H1 There will be no significant differences, immediately following the experimental treatments, between teachers of the Experimental Skills Group (ES), Experimental General Group (EG) and Control Group (C) on these variables:

1.1 frequency of asking various categories of questions,

1.2 proportion of questions asked in each category,

1.3 proportion of teacher to pupil talk,

1.4 level of Inquiry Score, and

1.5 patterns of teacher-pupil interaction.

In order to determine whether the three groups (ES, EG, and C) could be regarded as independent samples drawn from a single homogeneous population, a multivariate analysis of variance was first computed on the pretest data. Three separate analyses were carried out with the variables groups as follows:

(i) Frequency of Cognitive memory, Convergent, Evaluative, Divergent, Extension, Grounding, and Routine questions, and total frequency of questions.

(ii) Proportion of questions asked in the categories of Cognitive memory, Convergent, Evaluative, Divergent, Extension, Grounding, and Routine; the Inquiry Score; the proportion of teacher talk.

(iii) The proportion of the following interaction patterns:

a. Teacher speaks — one pupil speaks — teacher speaks.

b. Teacher speaks — two or three pupils speak — teacher speaks.

c. Teacher speaks — four or more pupils speak — teacher speaks and proportion of teacher to pupil talk.

The results of these MANOVA computations are presented in Tables 7.1, 7.2, and 7.3. No significant differences were revealed between the three groups of teachers (ES, EG, and C) on any of the three clusters of variables. Therefore, it could be assumed that each of the three groups of teachers consisted of a relatively representative sample of the total population from which they had been drawn, male Form One and Form Two Intermediate School teachers in Hamilton and Cambridge. On this basis, it was considered that, if significant differences were found in posttreatment data, these could be attributed to treatment effects, since the variables of time of lesson, objectives and content, and physical arrangements for recording lessons were controlled as far as possible.

Posttreatment teacher data were analysed independently of pre-

treatment data, i.e., the changes in performance between pre- and posttreatment were not examined directly, since their overall effect was not in question. For example, if all groups changed significantly in the same direction between two sets of observations, this would indicate only the effects of some common variable. Once the initial equivalence of the groups was established through the pretest, however, any subsequent significant differences (i.e., differential changes) could be attributed to treatment effects. By analysing the data in this way, any hidden regression effects could be excluded. Thus, where changes are referred to in the following discussion their significance is established by inference from the posttreatment differences rather than directly.

Computations on MANOVA for the above clusters for posttreatment data showed that there were significant differences between the treatment groups on each cluster. Table 7.1 (p.192) shows differences significant at the $p \leq .01$ level for frequency of questions asked in the first posttreatment lessons, and also at $p \leq .01$ for the delayed lessons. Table 7.2 indicates significant differences ($<.001$ and $<.05$ levels) for proportions of questions asked in each category, for the inquiry ratio and for teacher talk; while Table 7.3 reveals differences in teacher-pupil interaction patterns significant at the $p \leq .01$ level for both immediate and delayed posttreatment lessons.

After establishing that overall posttreatment differences existed, one-way analysis of variance was computed for each of the teacher variables. When the F ratio for any variable was significant at the $p \leq .05$ level, Duncan's New Multiple t Test was computed to find which of the three treatment group means were different from the others at either the $p \leq .05$ or $p \leq .01$ level of significance. In the statistical tables presented in Volume Two, the underlining method has been used to indicate the multiple t tests results. The group means have been arranged in ascending order of mean size from left to right using these symbols: ES = Experimental Skills Group; EG = Experimental General Group; C = Control Group. A line has been drawn beneath those groups which did <u>not</u> differ significantly at the level specified. Therefore those groups not joined by underlining, differed significantly at the level stated (Huck, Cormier and Bounds, 1974, p.70).

For reference purposes raw data are summarised in Tables 7.4 to

7.11, including the observed frequencies and derived ratios for each of the 24 teachers. As an additional check on the initial equivalance of the three treatment groups, separate one-way analyses of variance were carried out on the pretreatment data for each of the twenty teacher variables.

The results of these computations are shown in Tables 7.12 to 7.31 (pp.201-210). It will be noted that there were no significant differences, on any variable, between the means of the three teacher groups before the experimental treatment.

Hypothesis 1.1

Data and results of computations relevant to this hypothesis regarding frequency of use of various question types are summarised in Tables 7.32 to 7.39 (pp.211-214). Results from one-way analysis of variance computations indicated that group means differed significantly for these question categories: Cognitive memory (.001 level), Convergent (.01), and Divergent thinking (.001). The multiple-t tests (Table 7.32) showed that the frequency of Cognitive memory questions was significantly higher, on the average, for EG teachers than for either of the other two groups at the $p \leq .01$ level of significance. This resulted from an increased use of this type of question from pre- to posttreatment by the EG teachers, compared with either little change or decreasing use by teachers of the other two groups (Table 7.10).

This table also shows that the number of Convergent questions asked by teachers in the EG and C groups were similar in pre- and posttreatment lessons but the ES group number was halved in the post-treatment observations. ES teachers asked significantly fewer Convergent questions (.05 level) than either the EG or C teachers on the immediate posttest (Table 7.33).

Divergent thinking questions were asked at a mean frequency of less than two per lesson by each teacher group in the pretreatment lessons, but in the posttreatment lesson, whereas the EG and C groups stayed at that level, the ES group increased its mean to 4.9. On the posttreatment means the difference between groups was significant at the $p \leq .01$ level on Duncan's t test.

There were no significant differences between the immediate

posttreatment means of the three groups on the frequency of Evaluative questions (Table 7.34), Grounding (Table 7.36), Extension (Table 7.37), and Routine (Table 7.38). From Table 7.10, it can also be seen that the mean frequencies in these categories were fairly consistent between pre- and posttreatment lessons.

The overall frequency of questions asked by ES teachers dropped from a mean of 76.0 in the pretreatment lesson to 53.5 (Table 7.10); EG teachers increased their frequency slightly from 86.3 to 88.6; and C group teachers asked fewer (81.0 dropping to 71.9). Table 7.39 shows that ES teachers were significantly different from EG teachers (.05 level) on the posttest.

In line with the results presented above, the null hypothesis was rejected, as significant differences were found for three of the seven question types.

ES group teachers asked significantly fewer Convergent questions than the others, fewer Cognitive memory questions than the EG teachers, and more Divergent thinking questions than those in the other two groups. These posttreatment differences could reasonably be attributed to treatment effects.

Hypothesis 1.2

Data and results of computations relevant to this hypothesis related to the proportion of questions asked in each category are shown in Tables 7.40 to 7.46, and raw data for each teacher are presented in Tables 7.6 and 7.7.

The proportions were calculated in the following way from raw data for each teacher:

> Primary questions: frequency in each category as a proportion of total primary questions asked.
> Secondary questions (Extension, Grounding, Routine): frequency in each category as a proportion of all questions (primary and secondary).

For immediate posttreatment data, one-way analysis of variance for each of the seven question categories showed that there were significant differences between the ES, EG, and C groups for Cognitive memory questions (.001 level), Evaluative questions (.01), and

Divergent thinking questions (.001). A comparison of Tables 7.20 and 7.40 shows that the ES teachers substantially reduced their mean percentage of Cognitive memory questions in the posttreatment lessons, from 51% to 30% (rounded to nearest whole number), whereas proportions for both EG and C teachers increased. Duncan's t test confirmed that ES teachers differed significantly (.01 level) from EG and C teachers on the immediate posttreatment means.

The ES teachers increased the proportion of Evaluative questions they asked in the posttreatment lessons from a pretreatment level of 25% (Table 7.22) to 35% (Table 7.42), but the EG and C teachers reduced their proportions of this type of question. The multiple-t tests showed that ES teachers differed from EG and C teachers on the immediate posttest at the .01 level of significance (Table 7.42).

The proportion of Divergent thinking questions asked by ES teachers increased from a pretreatment mean of 2% (Table 7.23) to 15% in the posttreatment lessons (Table 7.43), with EG and C teachers asking similar percentages of these questions in both pre- and post-treatment lessons. Posttreatment group differences between ES teachers and EG and C teachers were significant at the $p \leq .01$ level (Table 7.43).

No significant differences were found between ES, EG and C teachers in the proportion of Convergent (Table 7.41), Grounding (Table 7.44), Extension (Table 7.45), and Routine questions (Table 7.46) asked in either pre- or posttreatment lessons. For each group, there was a high degree of consistency between the proportions of questions asked in each category in both pre- and posttreatment lessons.

The null hypothesis was therefore rejected, because in the immediate posttreatment lessons, ES teachers asked significantly higher proportions of Evaluative and Divergent questions, and a significantly lower proportion of Cognitive memory questions than both EG and C teachers.

Hypothesis 1.3

Data and results for this hypothesis regarding the proportion of teacher to pupil talk are summarised in Tables 7.8, 7.27, and 7.47. Raw data showed that there was a wide variation in the proportion of

total talk taken up by the teacher in the social studies lessons observed. Table 7.8 shows that the teacher talk in pretreatment lessons ranged from 40% to 84%, and the means for groups were: ES, 66%; EG, 58%; and C, 62%. In the immediate posttreatment teaching, the proportions were 40%, 66%, and 64% respectively thus indicating that while ES teachers — on average — decreased their proportion of talk by 26%, the EG and C teachers actually increased theirs slightly. The posttreatment F ratio was significant at the $p \leq .001$ level, and the Duncan's t test showed a $p \leq .01$ level of significant difference between ES teachers and both EG and C teachers (Table 7.47).

The null hypothesis was therefore rejected.

Hypothesis 1.4

The results of computations to test this hypothesis, which concerned the level of Inquiry Score, are given in Tables 7.9, 7.28 and 7.48. The Inquiry Score was computed by summing the frequency of use of those question types which were considered to be particularly appropriate for inquiry in social studies (Evaluative, Divergent, Extension, and Grounding), and calculating that score as a proportion of all questions asked.

For pretreatment data, Table 7.9 shows that within each experimental group, there was a wide range of Inquiry Scores: ES teachers ranged from .23 to .58; EG teachers from .21 to .57; and C teachers from .15 to .60. Ranges for immediate posttreatment lessons were as follows: ES teachers, .43 to .68; EG teachers, .16 to .40; and C teachers, .07 to .49. Put another way, the Inquiry Scores on the posttreatment data showed that all of the eight ES teachers were higher than the highest-scoring EG teacher, and only two of the C teachers had a higher Inquiry Score than the lowest score in the ES group.

Table 7.9 also shows that the pretreatment mean Inquiry Scores for each of the three experimental groups were similar — .36 for ES teachers, .37 for EG teachers, and .35 for C teachers. However, the immediate posttreatment mean for ES teachers rose to .55, whereas the mean for EG teachers was .28 and for C teachers, .33. The means for EG and C teachers were thus slightly lower in the immediate posttreatment lesson. The F ratio for differences in posttreatment group means was

significant at a level of $p < .001$ and the multiple-t tests indicated that the ES teachers' mean was significantly higher (.01 level) than the means for EG and C teachers.

On the basis of this evidence, the null hypothesis was again rejected.

Hypothesis 1.5

The computations and results relative to this hypothesis about patterns of teacher-pupil interaction are given in Tables 7.11, 7.29 to 7.31, and 7.49 to 7.51. The raw data for the teacher-pupil interaction variables were gathered from the transcripts of the lessons taught by the three sample groups of teachers, by counting the number of separate pupil utterances which followed each teacher utterance. From these, the frequencies of three basic interaction patterns were calculated.

The pattern 'teacher speaks — one pupil speaks — teacher speaks' made up over 90% of the interactions for ES, EG and C teachers in the pretreatment lessons. Individually, only two ES teachers' lessons, three EG teachers' lessons, and one C teacher's lesson had less than 90% of these patterns of interaction, the lowest being the 77% of an EG teacher (Table 7.11). Table 7.49 shows that in the posttreatment lesson immediately after the in-service courses, there was a significant difference (.01 level) between the proportion of these patterns in ES group lessons compared with those in EG and C group lessons. This difference resulted from ES teachers decreasing their use of these patterns from a mean of 93% (pretreatment) to a mean of 71% in the immediate posttreatment lesson, whereas the mean proportion for EG and C teachers remained at over 90%. In these posttreatment lessons, only one ES teacher's proportion of these interaction patterns was over 90%, while another ES teacher's lesson had only 53% of these patterns (Table 7.11).

For the interaction pattern 'teacher speaks — two or three pupils speak — teacher speaks', Table 7.50 shows that the posttreatment lesson means revealed significant differences between the teacher groups and the multiple-t tests indicated that ES teachers' lessons had significantly more (.01 level) of these patterns than did the lessons of the EG and C teachers. The difference arose because the ES

teachers changed from a pretreatment mean of 7% to 22% in the post-treatment lesson (Tables 7.30 to 7.50), but the EG and C teachers remained at about the same level in both lessons, i.e., at a mean of less than 10%.

Posttreatment data in Table 7.51 show that in the lessons of ES teachers, the proportion of use of the interaction pattern 'teacher talks — four or more pupils talk — teacher talks' increased from 0.3 to nearly 7% between pre- and posttreatment lessons. EG and C teachers' mean proportion of these patterns was less than 1% for both pre- and posttreatment data. One-way analysis of variance indicated an F ratio significant at the $p \leq .001$ level of probability for post-treatment proportions, and the multiple-t tests showed that the ES teachers' mean was significantly different from the means of the EG and C teachers (.01 level).

The null hypothesis was therefore rejected.

7.2 Delayed posttreatment effects

This group of hypotheses parallels those just discussed and may be stated as follows:

H2 There will be no significant differences between teachers in the ES, EG and C groups twelve weeks after the first posttreatment lesson, on the following variables:

2.1 frequency of asking various categories of questions,

2.2 proportion of questions asked in each category,

2.3 proportion of teacher-to-pupil talk,

2.4 level of Inquiry Score, and

2.5 patterns of teacher-pupil interaction.

Hypothesis 2.1

Again the raw data (frequencies) for each type of question asked (for all teachers) are shown in Tables 7.4 and 7.5, and the group means, standard deviations, and standard error of the means in Table 7.10. Analyses of variance for each question type are shown in Tables 7.52 to 7.58. The data for this group of hypotheses were based

on observations of seven ES teachers, eight EG teachers and seven C teachers. It should be noted that this is one less in the ES and C groups than in the previous analyses.

For frequency of asking questions, analysis of variance indicated that on four of the seven questions categories, there were statistically significant differences between group means on the delayed posttest: Cognitive memory, Convergent, Divergent, and Extension questions. Table 7.52 shows that the F ratio for Cognitive memory questions was significant at the $p \leq .01$ level, and the multiple-t tests revealed that ES teachers were significantly different from EG teachers at the $p \leq .01$ level but not significantly different from C teachers. The frequencies for the delayed lesson followed a similar pattern to the first posttreatment lesson, as can be seen in Table 7.10.

The F ratio in Table 7.53 for differences between groups in frequency of asking Convergent questions was significant (.05 level) and multiple-t tests showed that ES teachers were different from EG and C teachers at the $p \leq .05$ level of significance. Raw data in Table 7.10 show that the mean for all three groups increased, but the mean for ES teachers (8.1) was much lower than that for EG (18.5) or C (16.3) teachers; this led to the significant difference between group means.

The mean number of Divergent thinking questions asked by ES teachers, 4.29, was much higher than the number asked by EG (1.63) and C (0.57) teachers. The F ratio was significant at the $p \leq .05$ level, and the multiple-t tests showed that the ES group was significantly different (.05) from EG and C groups (Table 7.55). This again was a similar pattern to that revealed in the analysis of immediate posttreatment lesson data given earlier.

Results in Table 7.57 show that there was also a significant difference (.05 level) between group means for frequency of Extension questions. However, the difference was not in the direction predicted in the research hypothesis. C group teachers asked twice as many of these questions as ES and EG teachers, a difference not apparent in previous lessons (Table 7.10). The multiple-t tests confirmed the significant difference (.05) between C teachers and EG and ES teachers (Table 7.57).

There was also a significant difference (.05 level) between group means for the total frequency of all question types (Table 7.59). The multiple-t tests showed that EG teachers were significantly different from ES teachers (.05 level). However, the mean for the ES group (58.14) was considerably lower than that either for the EG (93.88) or C (75.57) groups, as had been the case in the immediate posttreatment lessons.

No significant F ratios were found for frequency of usage of the remaining question types: Evaluative (Table 7.54); Grounding (Table 7.56); and Routine (Table 7.58).

On the basis of the above results, the null hypothesis was rejected for the hypothesis as a whole.

Hypothesis 2.2

Results relating to this hypothesis regarding proportions of questions asked in each category are given in Tables 7.60 to 7.66. Analysis of variance indicated that the differences between means for the teacher groups were significant for four of the seven question types: Cognitive memory, Evaluative, Divergent thinking and Extension. Differences for the remaining three categories did not reach a significant level, even though, for Convergent questions, the ES group mean of 19.86 was markedly lower than that for EG, 25.25 and C, 32.86 (Table 7.61).

Table 7.60 shows that there was a significant difference (.01 level) between groups for the proportion of Cognitive memory questions asked, and the ES group mean of 33.86 was considerably lower than that for the EG (53.75) and C (49.14) groups. On the primary questions, ES teachers were therefore asking Cognitive memory questions on average just over one-third of the time while EG teachers were asking them over half the time and C teachers just under half the time. Duncan's tests confirmed that ES teachers were significantly different (.05 level) from EG and C teachers.

The results in Table 7.62 show that there was a significant F ratio (.01 level) for differences in the proportions of Evaluative questions asked. The ES group differed significantly from the EG and C groups. This result indicates the desirability of analysing proportions as well as frequencies, since Table 7.54 shows no

corresponding significant difference between group means based on frequencies alone. More power has been added to the analysis, for in effect, although there was a non-significant difference in the number of Evaluative questions asked by ES teachers, the proportion of the teachers' effort directed to encouraging evaluative thinking in their pupils was significantly greater (on average) in ES teachers' classrooms than in those of either EG or C teachers.

A much higher proportion of Divergent thinking questions was asked by ES teachers (mean 13% of primary categories) than either EG or C teachers (3% and 1% respectively). The F ratio was significant at the $p \leq .01$ level, and the EG group was significantly different (.05 level) from the EG and C groups on the multiple-t tests.

In addition, there was a significant difference at the $p \leq .05$ level between groups in the mean proportion of Extension questions asked, as is shown in Table 7.65. It should be noted, however, that the highest mean was that of the C teachers (16%), followed by ES (11%) and EG (7%). Duncan's test indicated that at the $p \leq .05$ level, the EG group differed significantly from the C group, but not from the ES group.

As the above results showed that on several of the categories there were significant differences between group means, the null hypothesis was again rejected.

Hypothesis 2.3

The proportion of teacher-to-pupil talk in the delayed post-treatment lessons, as shown in Table 7.8, was similar to that in the immediate posttreatment lessons, so that ES teachers talked, on average, just over one-third (36%) of the total talk-time (Table 7.67), while EG and C teachers took up a mean of two-thirds of the talk-time in their classes. The ES group was significantly different (.01 level) from the EG and C groups. In these delayed social studies lessons, ES teachers were able to reduce their proportion of classroom talk by a further 4%, on average, than in the immediate posttreatment lessons, the range of teacher talk being 16% to 55%, with only one teacher doing more than half the talking.

The null hypothesis was rejected.

Hypothesis 2.4

Table 7.68 shows the results of a one-way analysis of variance for the Inquiry Score for the delayed lessons. There was a significant difference between group means at the $p \leq .001$ level, and Duncan's test showed that ES teachers' scores were significantly higher than those for EG and C teachers at the $p \leq .05$ level. In effect, these results indicate that the combined question categories of Evaluative, Divergent thinking, Extension, and Grounding, were being used significantly more by ES teachers than by those in the other two groups.

The null hypothesis was again rejected.

Hypothesis 2.5

The results for the hypothesis related to the proportion of the interaction pattern 'teacher talks — one pupil talks — teacher talks' are given in Table 7.69, and indicate a difference between means at the $p \leq .001$ level of significance. Duncan's test shows that the ES group was significantly different (.01 level) from the EG and C groups. Proportions of use of this pattern were similar to those for immediate posttreatment lessons (Table 7.49) for each teacher group, showing that the ES teachers were able to maintain the reduced level previously shown.

The means for the pattern 'tacher talks — two or three pupils talk — teacher talks' (Table 7.70) were again similar to immediate posttreatment means (Table 7.50). The F ratio was significant at the $p \leq .01$ level, and the ES group mean differed from both the EG and C groups at the $p \leq .01$ level.

This consistency continued with the pattern of 'teacher talks — four or more pupils talk — teacher talks'. A mean of less than 1% of these patterns occurred in EG and C teachers' lessons, while the ES teachers' mean of 6% meant that they maintained a level close to the 7% they had reached in the other posttreatment lesson (Table 7.11). The F ratio was significant (.001 level) and the ES group differed from EG and C groups at the $p \leq .01$ level (Table 7.71).

The null hypothesis was therefore rejected.

7.3 Changes in teacher attitudes

The hypothesis here was that:

H3 There will be no significant differences between ES, EG, and C teachers in their posttreatment attitudes as measured by:

3.1 Rokeach Dogmatism Scale, Form E.

3.2 Zevin Attitude-to-Inquiry Scale.

Results of analysis of variance of pretreatment scores on the Rokeach Dogmatism Scale, Form E, are shown in Table 7.72. No significant differences between groups were found. Posttreatment results are summarised in Table 7.73, and again, the F ratio was not significant. The pre-post means were very similar for the ES group. The EG group mean increased by nearly three points and the C group mean decreased by nearly seven points, but in each case these differences were approximately equivalent to the estimated standard error of the mean.

On the Zevin scale, no significant differences were found between ES, EG and C groups for either pre- or posttreatment means (Tables 7.74 and 7.75). Raw scores showed an increase of 1.59 points for ES teachers, a decrease of 0.25 for EG teachers, and an increase of 2.19 for C teachers between pre- and posttesting, all within one standard error of measurement.

The null hypothesis was therefore accepted.

7.4 Immediate posttreatment pupil changes

The hypotheses here may be stated as follows:

H4 There will be no significant differences between the scores of pupils of the three experimental groups of teachers (ES, EG and C) in tests based on the immediate posttreatment lesson:

4.1 Cognitive memory subtest.

4.2 Higher-order thinking subtest.

4.3 Evaluative thinking subtest.

4.4 Divergent thinking subtest.

To test the above hypotheses analysis of covariance was used. Since the classes in question were the usual classes taught by the sample teachers, and therefore initial differences in ability probably existed, this statistical technique would take any such differences into account.

In order to maximise the effectiveness of the analysis of covariance procedure, a computer-assisted series of analyses was carried out to find the best covariates for each of the four subtests. Stepwise regression analysis was used to determine the relationship between age, intelligence quotient, PAT Reading Comprehension, pretest scores, and posttreatment scores. For the immediate posttreatment Cognitive memory subtest, the regression equation, Cognitive memory subtest = + .01 (I.Q.) + .28 (PAT Reading Comprehension) + .05 (Cognitive memory pretest), accounted for 17.79 of the variance.

Tables 7.76 and 7.77 give details of means, standard deviations and standard errors of the means for the pretreatment and posttreatment subtests respectively. Table 7.88 shows similar details for intelligence quotients, PAT Reading Comprehension scores, and age of pupils of the three groups (ES, EG and C) of teachers. Also, Table 7.93 shows Pearson product moment intercorrelations between all covariate data and subtest scores for the total pupil population used in this study.

The results of analysis of covariance and Duncan's New Multiple Range t Test are presented in Tables 7.79 to 7.82 for the four subtests. Table 7.79 indicates that for the Cognitive memory subtest, there was a significant difference between the group means at the $p < .05$ level, and Duncan's test shows that at the .05 level, ES group pupils scored significantly lower than C group pupils, but not significantly different from EG pupils.

The regression equation, Higher-order subtest = + .03 (I.Q.) + .67 (PAT Reading Comprehension) + .27 (Higher-order thinking subtest, pretest scores), accounted for 47.20 per cent of the variance of

immediate posttreatment Higher-order thinking subtest scores.

Results for the Higher-order thinking subtest scores are shown in Tables 7.77 and 7.80, and the analysis of covariance revealed no significant differences between the means of pupils of ES, EG and C teachers.

The regression equation, Evaluative thinking subtest scores = + .07 (age) + .14 (I.Q.) - .37 (PAT Reading Comprehension), accounted for 9.83 per cent of the pretest variance apart from Evaluative thinking pretest scores. More sensitive results were obtained by using the pretest Evaluative thinking subtest scores alone as the covariate and the immediate posttreatment test as the criterion variable. The regression equation, Evaluative thinking subtest = + .78 (Evaluative thinking pretest scores), accounted for 39.61 per cent of the variance of the scores for the immediate posttreatment Evaluative thinking subtest. Therefore the pretest scores were adopted as the covariate for analysis of covariance of these posttreatment scores.

Stepwise regression analyses to find the best covariates for analysing posttreatment Divergent thinking scores, indicated that the regression equation, Divergent thinking subtest scores = + .71 (Divergent thinking pretest scores), accounted for 43.56 per cent of the variance for immediate posttreatment Divergent thinking scores.

Table 7.82 shows that analysis of covariance using pretest scores as the covariate, revealed a significant difference between group means at the $p < .001$ level for Divergent thinking subtest scores. Duncan's tests showed that ES pupils again scored significantly higher than the EG and C pupils at the $p < .01$ level.

Summarising the results for pupil attainment on the immediate posttreatment test, there were significant differences between group means for the Cognitive memory, Evaluative thinking, and Divergent thinking subtest means, but no significant differences for Higher order thinking subtest means.

The overall null hypothesis was consequently rejected.

7.5 Delayed posttreatment pupil changes

The hypotheses, parallel to those for the immediate post-treatment tests, may be stated as follows:

H5 There will be no significant differences between the scores of pupils of the three experimental groups of teachers (ES, EG and C) in tests based on the lessons taught twelve weeks after the first posttreatment lesson:

5.1 Cognitive memory subtest.

5.2 Higher-order thinking subtest.

5.3 Evaluative thinking subtest.

5.4 Divergent thinking subtest.

Results for these hypotheses are presented in Tables 7.78, 7.83, 7.84, 7.85 and 7.86. For each of the four subtest scores, stepwise regression analysis was again used to find whether the most effective covariates to be included in analysis of covariance were those used in analysing the immediate posttreatment data. Analyses revealed that this was, in fact, the case: the same covariates accounted for the greatest amount of variance in the respective subtest scores.

The regression equation, Cognitive memory subtest scores (delayed) = + .01 (I.Q.) + .29 (PAT Reading Comprehension) + .07 (Cognitive memory pretest scores), accounted for 18.30 per cent of the variance of the delayed posttreatment scores for the Cognitive memory subtest. Table 7.83 shows that there was no significant difference between group means.

The regression equation, Higher-order thinking subtest scores (delayed) = + .04 (I.Q.) + .26 (PAT Reading Comprehension) + .26 (Higher-order thinking pre-test scores), accounted for 38.00 per cent of the total variance for posttreatment (delayed) scores for the Higher-order thinking subtest). Analysis of covariance using these covariates indicated that there were no significant differences between pupils of ES, EG and C teachers (Table 7.84).

Through stepwise regression analysis, it was again found that pretest scores were the best predictors of delayed-test performance for Evaluative thinking. The regression equation, Evaluative thinking

subtest (delayed) scores = + .54 (Evaluative subtest scores, pretest), accounted for 29.73 per cent of the variance. Table 7.85 shows that there was a significant difference between group means at the $p \leq .001$ level, and Duncan's test indicated that the ES group mean was significantly higher than those from the EG and C groups at a $p \leq .01$ level.

The regression equation, Divergent thinking subtest scores (delayed test) = + .68 (Divergent thinking pretest scores), accounted for 37.79 per cent of the variance of posttreatment (delayed) scores for the Divergent thinking subtest. Results of analysis of covariance (Table 7.86) reveal a $p \leq .001$ level of significance for differences between group means, and the Duncan test again shows that ES group pupils performed significantly better (.01 level) than EG and C group pupils.

In summary, then, there were significant differences between group means on two of the four subtests — Evaluative thinking and Divergent thinking — and no significant differences on Cognitive memory and Higher-order thinking subtests.

The null hypothesis was rejected.

7.6 Changes in pupil inference abilities

Stated in the null form the hypothesis here was that:

H6 There will be no significant differences between pupils of the three experimental groups of teachers (ES, EG and C) on Inference and Caution scores of the *Social Studies Inference Test*, either before or after the experimental treatment.

Table 7.87 summarises the pretest and posttest means, standard deviations, and standard errors of the means, for Inference and Caution scores. The means varied little between the three groups of pupils, and between pre- and posttests. Analysis of variance of the pretest scores on the one hand (Table 7.89) and posttest scores on the other (Table 7.90), showed that there were no significant differences between ES, EG and C pupils.

The null hypothesis was therefore accepted.

7.7 Summary

Two major groups of variables were the subject of experimentation in this study. One group related primarily to teacher behaviours, and the other group to pupil behaviours. For teacher variables, multiple analysis of variance was used to determine whether any significant differences existed between experimental group means on the pretreatment measures, and none were found. Consequent multiple analysis of variance of posttreatment data indicated that there were significant differences between the group means for each of the three clusters of variables. To find exactly where these differences were, one-way analysis of variance was used for each teacher variable in turn, and where significant F ratios were found, Duncan's New Multiple Range t test was used to find which groups differed significantly.

As a result of the analyses of data on teacher questioning, for both the immediate and delayed posttreatment observations, the null hypotheses were rejected overall. However, closer scrutiny of the results shows that while some of the specific categories, such as the frequency with which teachers used different kinds of questions, showed significant posttreatment differences between the teacher groups, others did not. No significant differences were found, either pre- or posttreatment, in teacher attitudes.

Pupil attainment formed the second major group of variables analysed. Analysis of covariance was used to weight group means to compensate for initial differences which might have existed between the experimental group pupils. On the basis of analyses of immediate posttreatment data, hypotheses of no difference in pupil attainment on Cognitive memory, Evaluative thinking and Divergent thinking were rejected, while for the delayed posttreatment data, there were significant differences between pupils only on Evaluative thinking and Divergent thinking. There were no significant differences either pre- or posttreatment, on pupil Inference and Caution test scores.

The statistical analyses therefore revealed a number of significant posttreatment differences in both teacher and pupil

variables which could be attributed, directly or indirectly, to the experimental in-service programmes. These treatment effects will be discussed in some detail in the following chapter.

CHAPTER 8

DISCUSSION

OVERVIEW: In this chapter the results which were presented in the previous chapter are discussed in some detail. Comparisons are made between baseline data in this study and in other studies referred to earlier. Certain changes in teachers' questioning resulted from the specific-skills in-service programme, along with other verbal behaviour changes, and these are discussed. Some apparent effects of these changes on the inquiry climate in the classroom are described, and the extent to which changes in verbal behaviour were maintained over a three month period are examined. The posttreatment test scores of ES group pupils for evaluative thinking and divergent thinking differed significantly from the scores of the EG and C group pupils, and the possible reasons for these changes are discussed.

The data collected through taperecordings of the pretreatment lesson taught by each of the teachers in the sampe provided 'baseline' information about the verbal behaviour of 24 male Intermediate School teachers in social studies in the present study. There has been a paucity of such data gathered in New Zealand settings.

It is interesting to compare this baseline data with that from other studies in different situations. As Tisher (1970) found in science lessons, and Gallagher, Aschner and Jenné (1967) found in social studies, although the patterns of questioning vary between individual teachers, there are reasonably consistent questioning patterns that each teacher has developed, especially if the lesson material is similar in each lesson. In the present study, the pretreatment lessons taken by the 24 teachers showed that the proportion of questioning behaviour in the four primary categories ranged from 30 to 76%. The use of Convergent thinking questions ranged from 6 to 44%; Evaluative from 7 to 57%, and Divergent thinking questions from 0 to 7% (Table 7.3). Fifteen of the 24 teachers asked no Divergent thinking questions at all in these pretreatment lessons.

Gallagher, Aschner and Jenné (1967) found a similar pattern in social studies lessons with gifted seventh and eighth grade pupils; that is, a predominant use of Cognitive memory questions, and few, if any, Divergent thinking questions. However, in the present study, the New Zealand teachers asked a higher proportion of Evaluative questions than those in the Gallagher, Aschner and Jenné study; most teachers in the United States sample asked less than 10% of this type of question, but in the present study ES teachers asked a mean proportion of 24.63%, EG teachers 21.63%, and C asked 22.75% in pretreatment or baseline lessons. Table 7.6 shows the individual teacher's proportions.

Aschner, Gallagher and Jenné (1967) also found that in different class sessions, the same teacher in social studies might ask different proportions of each question category. For example, in a series of five lessons taught by one teacher, 40% of his questions in the second lesson were Cognitive memory, compared with 70% in the fifth lesson. Another teacher asked over 70% Cognitive memory questions in all five lessons of one series, then asked fewer questions of that type in three lessons of another series (*ibid.*, pp.38-9). However, this observed intra-individual variability may well have resulted from the researchers not standardising the lesson content.

On the basis of the above, and the Smith and Meux (1962) findings, and because Gall (1970) and others had strongly recommended it, the content of the three lessons taught by each teacher in the present study, was standardised by the investigator. It was assumed that this would lessen the extraneous variation that Smith and Meux had noted, and would enable differences between ES, EG, and C groups in the frequencies and proportions of questioning to be more reliably attributed to treatment effects.

8.1 Changes in teacher questioning behaviour

The results presented in the previous chapter indicate that the in-service course which focussed upon specific skills of questioning did make significant differences to the questioning behaviour of the ES teachers. The general hypotheses relating to the effects of the two in-service courses predicted changes in the subsequent

questioning behaviour of teachers in the ES group but not necessarily in the behaviour of those in the other two groups. Results of the testing of the null hypotheses, however, showed that the teachers who were given an in-service course that treated inquiry and questioning in a more general way (the EG group) did not change their questioning behaviour significantly as compared with the control group teachers.

Comparisons were made between the ES, EG, and C teachers on 15 variables directly related to teacher questioning. These variables were: frequency of asking seven question categories and total frequency of questions, and the proportion of questions asked in each of the seven categories. In summary, for the immediate posttreatment lessons, analysis of variance showed that the ES teachers were significantly different from the EG and C teachers on six of the variables, and the EG teachers were significantly different from the C teachers on one variable.

In the lessons taught after a three-month delay, the ES teachers differed significantly from both the EG and C teachers on seven of the variables, while the EG teachers differed significantly from the C teachers on one variable only.

More specifically, during both the SPRITE course for ES teachers and the more general one for EG teachers, it was suggested to them that there was merit in reducing the proportions of Cognitive memory, Convergent, and Routine questions, and in increasing the proportions of Evaluative, Divergent, Extension and Grounding questions in social studies discussion lessons which tried to foster inquiry learning. In the lessons taken shortly after both in-service courses, the ES teachers who took the specific-skills (SPRITE) programme, asked significantly fewer Cognitive memory questions (Table 7.32) and significantly more Divergent thinking questions (Table 7.35) than both the teachers who took the general course, or those in the control group.

As well as reducing their proportion of Cognitive memory questions, there was a dramatic change in the ES teachers' frequency of asking Cognitive memory questions, as a result of the SPRITE (ES) programme. In the pretreatment lessons the eight ES teachers asked a total of 219 Cognitive memory questions between them, EG teachers asked 265, and C teachers asked 187. However, in the immediate post-

treatment lessons, the same eight ES teachers asked only 83 Cognitive memory questions, yet the eight EG teachers asked an even higher number, collectively, than they had previously, 332, and the C teachers asked 199. These raw data show that, through the SPRITE programme, there had apparently been little difficulty in convincing teachers that there was merit in trying to substantially reduce the number of Cognitive memory questions they asked. Surprisingly, however, the stress on inquiry teaching in a general way, in the course for EG teachers, did not lead those teachers to reduce their use of the same recall questions, even though questioning skills were often referred to, and a reading about questioning was issued. The proportion of Convergent questions asked by ES teachers in their immediate posttreatment lesson did not differ significantly from the EG and C teachers, but the ES teachers' usage (frequency) of Convergent questions did differ significantly from that of the other two groups (Tables 7.41 and 7.33).

On the other hand, the frequency with which the ES teachers asked Evaluative questions increased and was significantly different in the immediate posttreatment lesson when compared with that of both the EG and the C teachers. Again there were no significant differences between groups in the proportion of Evaluative to all primary questions. During the SPRITE programme, it was noticeable that these New Zealand teachers were accustomed to asking Evaluative questions, and it was noted earlier that their pretreatment proportions of this kind of question were higher than reported in Gallagher's (Gallagher, Aschner and Jenné, 1967) American study. However, in the pretreatment lessons of all three treatment group teachers, many of the Evaluative questions tended to be closed; that is, they asked for simple 'yes' or 'no', or 'agree' or 'disagree' answers.

The posttreatment tapes indicated that, during the SPRITE course, the ES teachers learnt to ask Evaluative questions that were more open-ended and resulted in more extended responses from the pupils. For example, the following are typical Evaluative questions from the pretreatment transcript of one ES teacher:

T	Cooperation. Did that come through in this film?
P	Yes.
T	All right.

and,

T	Is it a comfortable job to do?
P	No.
T	Jillian.
P	You've got to make sure you don't take too long about it.
T	Right, 'cos otherwise the stones get cold.

In the lesson the same teacher taught shortly after the in-service programme, the following was more typical of the Evaluative questions:

T	Why was the trading post important in the lives of the Hopi Indians? David.
P	Because there they can go and trade their stuff and get food, clothes and medicines, and things that they needed.
P	It was their main source of food.
P	Well, when the white man arrived he probably killed all the animals and that they relied on for food and all that stuff, the clothes and that, so they have... have to have trading posts to get those things.
P	They'd keep up with technology.

The difference (pre- to posttreatment) seemed to be that in the second lesson, the teacher was trying to get the children to give their own opinions about a value question, and if possible to use evidence from their background knowledge which had been provided by the television programme they had seen. This was also observed in the case of other ES group teachers. It seemed that much more was happening than just a reduction in teacher talk. The pupils appeared to have more opportunity to reason and to delve into problems rather than quickly pass over them after having made a few superficial comments.

As a consequence of what has been said so far, it is not surprising that the ES teachers significantly reduced the number of questions they asked in the twenty minutes social studies lessons after the experimental treatment, from a pretreatment mean of 76.00 to a posttreatment mean of 53.50; or, stated another way, the three teacher groups (ES, EG, and C) asked an average of one question every 16 seconds in the baseline lessons, but whereas the ES teachers asked one question every 23 seconds in the immediate posttreatment lessons, the

other two groups maintained the 16 second rate.

The amount of lesson talk taken up by teachers decreased sharply for ES teachers after the experimental treatments. ES teachers apparently encouraged their pupils to talk more in post-treatment lessons, whereas EG and C teachers did not. In fact, EG teachers actually increased their proportion of talk, despite the discussions about inquiry methods during their in-service course. As suggested above, it seemed that by asking a higher proportion of Divergent and Evaluative questions (in the delayed lessons) the ES teachers were providing their pupils with the chance to give more elaborate answers, and were, at the same time, reducing their own dominance.

8.2 Changes in interaction patterns

The analysis of teacher-pupil interaction patterns illustrated this development even more clearly. The dominant pattern in all pretreatment lessons was 'teacher talks — one pupil talks — teacher talks'. A similar finding was made by Smith and Meux (1962) who, however, used a complex system of recording and analysing episodes. They found that the dominant episode in the classrooms they observed was 'teacher seeks — child gives — teacher comments'. The method of anlysing the interaction patterns in the present study was somewhat more simplistic, but nevertheless the three patterns identified did indicate changes in the extent to which teachers involved pupils in discussion.

In posttreatment lessons, the ES teachers made a considerable change, in that they encouraged pupils to give longer, more elaborate answers, and to add to each others' answers without waiting for (or being subjected to!) teacher intervention. It could be argued that these teachers were promoting discussion and inquiry to a greater extent than they had been in pretreatment teaching of a similar topic.

The following extract from a transcript illustrates this point. The excerpt is from a posttreatment lesson (Hopi Indians) taken by a teacher whose pretreatment lesson contained no sequence in which more than three pupils spoke in succession. In his first posttreatment

lesson, however, eighteen percent of the verbal patterns were of the most extended kind (four or more pupils speak in succession);

T How might they (Hopi Indians) receive a permanent water supply? Neil.

P A well.

P Well, in this film I was watching you just ... they just dug up all the sand, ... and finally they came to the wet sand and the water just came up and they got to it then.

P They could save all the water from the rain, if possible.

T How might they do that?

P Well, they might have great saucepans or something and they put all the water in that and they put it in a fairly protected place.

P Umm ... in this film [the television presentation] it said they had to carry the water to the, ah, peach trees to water them.

P Well, they had to get it from somewhere!

P Well, what Mary said can't be right because the climate ... it said the climate was a dry one, so it wouldn't rain much.

P Well, I disagree with Judy because ... um ... ah, sure it would be dry all the time, but there must ... at some time or other, there must be some rain.

P I don't think the, ah, Indians would rely on that because there's only a hundred and eighty millimetres a year ... of rain a year, and they couldn't really rely on it to be constant, as I they'd have about four or five different wells around the place, 'cos what would they catch the rain with anyway, and whatever rain they caught would be hard to store.

P Well, if they catch the rain when it comes, well, it would evaporate because it's so hot.

T Well, how do you imagine this rain would come when it does come? Keith.

P Well in the desert when it's really hot when it does rain, oh, it's just a huge downpour. It would be pretty muggy inside wouldn't it? It's just ... just like a huge, ah, storm. We might have it twice as bad.

P Ah, I believe too, that it rains a lot in the deserts and a lot of it dries up before it hits the ground. It rains ... it's dry before it gets down to the village.

P Well, if there's any cactus around, they might get water from that.

The above transcript illustrates clearly how this teacher encouraged the pupils to talk freely, in an inquiring manner. Two Divergent questions allowed the pupils to explore what they knew about the Hopis, to bring previous knowledge about the desert into the discussion, along with information they got from the lesson introduction. The teacher was content to allow pauses for children to re-think an answer, or to change the emphasis in the middle of giving an answer. When one child made a contribution, the teacher waited (this was evident from the taperecording) to see whether another child wished to participate. It should also be noted that the children felt freer to disagree with one another than they seemed to be in the pre-treatment lesson. What has been said above is in no way intended to suggest that the amount of pupil talk equates with the level of their inquiry. It is a contributing factor.

This can be compared with an excerpt from an EG Group teacher in a lesson based on the same content:

T Now, it said in the film also that they live in a pueblo. What's a pueblo?

P Their ... their, ah, religion.

T No.

P A home.

T Well, it's a group of homes. What do you call a place where a group of homes are built? It's Spanish for something.

P A village?

T Right, it's Spanish for village. That's something I've learnt from the film too. I thought pueblo was a particular band of Indian, but apparently it's a village. O.K. What was this Spanish word for table that Bruce was trying to tell us about before? It wasn't maize was it? That's like corn. Can you remember?

P Mesa.

T Mesa, wasn't it. Right. What was ... did you find particular characteristics about the mesa? What does it look like? What was it?

P A top of a table.

T That's what it meant, wasn't it, a table top. Well, what did one look like? Verna.

P A hill with a flat top.

T Well, what sort of hill with a flat top? Just an ordinary hill with a flat top?

P Small jagged hills.

P Cliff with a flat top.

T Who thinks he's closest so far? [some hands are raised]. Well, carry on with the description, Ross.

P Rocky hills, with cliffs and a flat top.

T Yes. What about the sides? I suppose cliff describes that, yes. Mind you I suppose they had to have sort of tracks to get down, didn't they? They didn't have a ski lift or anything like that to get down did they? They'd have to have a path up. Why did they build on top of the mesa? For what reason would you say, Brenda?

P To keep away their enemies or the other tribes.

T Right. In one word then, why were they there then? One word.

P Safety.

T Right. Why is the weather so important for these Indians?

P Because of their crops.

T All right. Who can tell me something about their crops? What crops do they grow?

P Corn.

This excerpt contrasts markedly with the previous one, in that the teacher here is much more dominating in amount of talk. He also asks a series of memory recall questions for which the children try to give the answer he seeks. The answers to all of the questions were provided in the introductory television material — either by being said directly, or by being shown visually. Even where a description of a mesa was asked for, it seemed that the teacher had a preconceived set of answers, which he wanted to bring out and yet here was an ideal chance for the children to give their own impressions. It is important to note that although this teacher used the terms 'for what reason' and 'why' to preface questions, they were in fact, no more than recall questions because the information asked for had been given in the lesson. This points up the necessity — when coding questions — to know the full context of the lesson in which they are asked. What might appear to be a question asking for higher-level thinking, might, in fact, not be so. In this respect, written transcripts would seem

much better for coding then trying to code as a lesson progresses as in some systems.

8.3 A climate for inquiry

It was also noted by the investigator, both when taperecording the posttreatment sessions and when listening to the tapes afterwards, that the ES teachers developed an increased willingness to wait for children to formulate answers. This technique appeared to be poorly developed in all 24 pretreatment lessons, and although this could possibly have been the result of teacher nervousness in the initial observation session, it was noticeable that the posttreatment lessons of the EG and C teachers were similar to their pretreatment ones in this respect. Teachers seemed to be unused to letting silence continue for more than a second or two. In the posttreatment lessons, the children of ES tachers seemed to be aware that the teacher was prepared to wait until some thought had been given about the answer. It should also be noted that the skills in-service course (SPRITE) had included emphasis upon this technique. However, no systematic analysis of wait-time was carried out, so the above comments are subjective only.

Nevertheless, it seemed clear that, in their posttreatment lessons, the ES teachers were more willing to wait after the first answer to an open question, in case another child wanted to continue. The asking of a greater proportion of Evaluative and Divergent questions helped to facilitate this more flexible pattern of interaction, because ES pupils were encouraged to give their own ideas and opinions without being afraid of giving the wrong answer. In the posttreatment lessons of ES teachers the results presented previously show that there were many more extended patterns of talk in which pupils had 'runs' of comments or answers. Many writers (for example, Massialas and Cox 1966; Massialas and Sprague, 1974) have pointed to the necessity for pupil verbal involvement if inquiry thinking is to take place in social studies discussion lessons.

Furthermore, the significant increases which were shown by ES teachers on the Inquiry Score (ratio) indicated that they were able to alter the distribution of their use of the various question types. In

effect, an increase in Inquiry Score for the posttreatment lesson represented more Evaluative, Divergent, Extension and Grounding questions (in relation to Cognitive memory, Convergent, and Routine questions) than in the pretreatment teaching. The results indicated that it was not just a simple matter of the ES teachers talking less. There was a definite shift in the balance of the talk, which was initiated by a change in the kinds of questions asked. It can be argued that this represented a change in the quality of questioning, not just in its quantity. In other words, the increased proportion of pupil talk was, also, more than just a quantitative change.

The above behaviour changes seem to exemplify the adoption of new strategies for the ES teachers, and, at the same time, alterations in the classroom climate. The 'intellectual permissiveness' that Postman and Weingartner (1969) and Marsh (1975) saw as important, seemed much closer to being realised in that pupils were freer to contribute than in pretreatment lessons. However, it must be noted that such changes did not occur evenly over all ES teachers (Tables 7.4 to 7.10).

8.4 Were the changed behaviours maintained?

One of the inadequacies of much classroom research noted by reviewers (for example, Rosenshine, 1971 and Borg, 1975) is that most studies have collected data over a very short period of time, and there has therefore been little information about how lasting are the effects of changes in teacher behaviour brought about by in-service courses. In an attempt to find whether changes did last longer than immediately after the in service courses conducted in this study, a lesson was taperecorded 12 weeks after the first posttreatment lesson; that is, about three months after the in-service courses had been held.

The results presented earlier showed that, in the first post-treatment lesson, the EG teachers made few changes in their pre-treatment behaviour on the variables analysed. The ES teachers did, however, change their behaviour in several relevant respects, making them significantly different from teachers in the other two groups, and these changes were attributed to the effects of the in-service

treatment. In the delayed lesson, results showed that the ES teachers were able to maintain the magnitude of these changes to a large extent, and, in some cases to increase them.

In both posttreatment lessons, ES teachers' patterns of questioning were very similar (Table 7.4). The significant reductions in the numbers and proportion of Cognitive memory and Convergent questions were maintained in both immediate and delayed posttreatment lessons. The frequency and proportion of Divergent thinking questions remained at a higher level in the ES teachers' delayed lessons, than in their pretreatment lessons, and the proportion of Evaluative thinking questions they asked also remained higher, and these represented significant differences from the questioning behaviour of the remaining teachers. One ES teacher increased his Cognitive memory questions to more than his pre-treatment total, yet also increased (14 to 23) Evaluative questions. Another asked no Divergent questions in the delayed lesson, but had asked five in the first posttreatment teaching. Such idiosyncratic behaviour must always be expected, especially where the content matter of lessons has changed.

It might also have been expected, that as distance in time away from the in-service courses increased, the proportion of teacher-to-pupil talk would rise again, a tendency which Borg (1975) found in his minicourse teachers (though over a longer time — three years). However, somewhat surprisingly, the ES teachers actually reduced their mean proportion of total talk in the delayed lesson even more than they had managed in the immediate posttreatment lesson, so that they spoke for just over one-third of the talk time (Table 7.8). In the delayed lesson, one teacher spoke for only 17% of the total talk time whereas in the pretreatment lesson he did 60% of the talking. The same teacher had asked only 3% Divergent thinking questions, and 35% Evaluative thinking questions in his pretreatment lessons, and yet was able to ask 28% Divergent thinking and 40% Evaluative thinking questions in his delayed lesson. This indicates a probable relationship between a teacher's ability to increase his more open-ended questions, and to reduce his need to talk so much.

Of the other ES teachers, three talked more, and three less, in the delayed lesson than in the immediate posttreatment session. It

was noticeable that the greatest fluctuation in the proportion of talk among individual EG and C teachers in all three lessons, was 26%, with a median of 10% (Table 7.8), and overall, there was a high degree of consistency between all lessons for each teacher in these two groups. Expressed as means, the consistency was extremely high: 58%, 66%, and 66% for the EG teachers, and 62%, 64%, and 65% for the C group teachers.

In view of the findings referred to above, it was not surprising that the mean Inquiry Score for ES teachers was similar in both post-treatment lessons. Only one of these teachers' scores fell markedly in the delayed lesson — from .68 to .38 (Table 7.9). The Inquiry Score is thus considered to be useful as an exploratory device for judging, quantitatively, whether changes towards inquiry-type questioning had occurred.

8.5 Focussing on specific teacher skills

The results already presented have indicated that it is possible to give teachers an in-service course with a specific skills focus (such as SPRITE) which will enable them to translate some of the skills they have been taught into their classroom practice. These findings support studies by Taba (1966), Galassi *et al.* (1974), Merwin and Schneider (1974) and others in which in-service courses were successful in changing specific verbal behaviours such as questioning.

One of the main differences between the SPRITE course and the general programme, was that teachers in the former were taught to identify and practice specific questioning skills. Although the same skills were referred to in the general course, however, and had been supported by available readings about inquiry and questioning, with verbal affirmation of the value of questioning given by the teachers concerned, no translation into practice occurred in their case. This gives support to the contentions of Bush (1973) and Borg (1975), who have been critical of in-service courses with a general focus because they do not produce changes in teacher skills. It would seem, then, that if teachers' specific questioning skills are to be altered, the skills must be clearly defined and taught. It is not sufficient

for educators to assume that general reference to them through lectures and/or readings, will lead to changes in teachers' classroom behaviour.

Some recent programmes in in-service education have used videotape techniques for showing teachers their own performance, and given feedback about that practice (for example, Borg *et al.*,1970; Galassi *et al.*, 1974). However, in the SPRITE course written transcripts of examples of the skills were used, plus teacher taperecordings of their own teaching for feedback. Results indicate that this method was successful in helping to teach specific skills. Galassi *et al.* (1974) had used transcripts of skills, and found them at least as effective, if not more so, than videotape. It is possible that in a relatively short course like SPRITE, the novelty effect of television might actually inhibit or delay a concentrated effort upon the skills.

In both the SPRITE course and the general one, there were three days between the first two days of the course and the third and fourth days. In this time, each teacher taught his usual class. This provided the chance for both experimental groups to practice the techniques concerned. ES and EG teachers taught lessons and taperecorded them. ES teachers then got feedback from other course members by playing their tapes and hearing what the others had to say about the question types used in their teaching. EG teachers concentrated more on the effectiveness for inquiry learning of lesson materials that they had constructed in the first two days of the in-service course.

In these feedback sessions, the ES teachers concentrated upon questioning skills, almost exclusively, but the EG teachers talked about questioning as simply a component of inquiry teaching. It could well be, then, that the effectiveness of in-service courses is increased if, at first, a narrow range of skills is concentrated upon, rather than a number of skill clusters that relate to a somewhat vague teaching method like 'inquiry'. The present research also gives support to the claim (Lawton and Dufour, 1973) that new skills, such as teacher questioning in inquiry-based social studies lessons, are hard for teachers to learn since, even with concentrated practice, some ES teachers found particular questioning skills difficult to master.

The asking of Divergent questions proved to be the most challenging aspect of the SPRITE programme for ES teachers. In the eight pretreatment lessons of the ES group, only 10 Divergent questions were asked, and yet the lesson content was suitable for many more of such questions; three ES teachers asked no Divergent questions at all. Closer scrutiny showed that for most of the ten questions, the teachers concerned had the answers fairly clearly in their minds, and the discussion that followed each question was quite limited. During the in-service programme, lengthy discussions were held about whether pupils would learn through attempting to answer Divergent questions. Some teachers at first expressed the opinion that unless the children were asked numerous Cognitive memory questions early in the lesson, there was no guarantee that they had learnt a body of information which they could then use as a basis for being divergent or creative, or giving their own opinions. The researcher argued that the children had already been presented with information, and that in answering Divergent questions they would use it, and by listening carefully to pupils' answers, teachers would soon know if they had in fact remembered the information. The teachers were asked to think about this in lessons they took in their own classes after the first two days of the SPRITE course. It was noticeable that, in these practice lessons, all eight teachers tried to ask Divergent questions reasonably early in the lesson. Each taperecording was listened to by the eight teachers, many comments were made, and a most interesting discussion resulted. In other words, what seemed to be happening was that the teachers were realising that there was no need to ask as many Cognitive memory questions as they had been doing previously, because the answers which pupils gave to the Divergent questions were not only more thoughtful, but showed that they had remembered more than the teachers thought they had. Bruner (1961) and Suchman (1962) had been confident that the growth of knowledge would 'follow in the wake of inquiry'.

Another important factor in relation to Divergent questions was that ES teachers said they did not ask many of them because they had never taken 'time out' to give concentrated attention to them. Two teachers, for example, were sceptical at the start of the SPRITE programme, that teachers should ask questions that they themselves did

not know the answers to. By the end of the course, both had changed that viewpoint and their posttreatment lessons substantiated this. The taperecordings of the posttreatment lessons supplied considerable evidence that the ES teachers had indeed mastered new questioning techniques (such as asking more Divergent questions) which enabled them to reduce what Suchman (1961) called their 'traditionally directive mode', and to encourage more freedom of expression on the part of pupils.

It is interesting that in informal conversation with teachers at the end of the data collection phase, each ES teacher reported that he had attempted to work on the further development of his own questioning strategies, such as asking more open questions. The EG teachers generally said, however, that they were interested in developing inquiry techniques, but they tried to do so by constructing materials rather than by changing their own specific questioning skills. Their comparative lack of success in encouraging inquiry thinking in their pupils is reflected in the analysis of the interaction patterns referred to earlier.

8.6 The lesson context

Taba (1966) noted that when educators try to teach skills through in-service courses, the behaviours should be taught within a particular context, such as a subject field. The present research focussed upon cross-cultural topics in social studies; the three packages were all village studies, based upon a common set of social science concepts. This meant that the in-service courses provided opportunities for discussion about inquiry methods and questioning which could involve examples from materials familiar to all the teachers. Additionally, it was assumed that similar patterns of questioning were appropriate for each package.

It was noted in Chapter 2 that curiosity arousal is an important role of the teacher in an inquiry-based discussion (Day and Berlyne, 1971). The comparative success of the ES teachers in increasing the evaluative and divergent thinking behaviour of their pupils in the attainment tests, might have been the result of such arousal of their

pupils, because the teachers were able to ask significantly more Evaluative and Divergent questions; this, in turn, may have encouraged the pupils to talk (and think) more. The sharp increase in extended patterns of interaction observed in ES teachers' post-treatment lessons gives support to this claim. There were far more instances of several pupils carrying on what another had said, rather than automatically waiting for a teacher response to every pupil utterance. Pupils were given the chance to be more curious and thoughtful; whether they were more curious, however, cannot be demonstrated from this study. Associated with this curiosity arousal, it could be expected that pupils themselves would ask more questions in discussion, but there was not time to pursue this aspect further.

8.7 Teacher attitudes

Teacher attitudes — as measured by the *Rokeach Dogmatism Scale, Form E* and the *Zevin Attitude to Inquiry* scale — changed very little from pretest to posttest, and there were no significant differences between the experimental groups. Scores on the dogmatism scale for all teacher groups were slightly higher than those reported in Chapter 5 for American University students, but lower than those for English adult 'workers'. The scores for individual teachers did not reveal, from inspection, any evidence of Flanders' (1970) claim that high-dogmatic teachers are less likely to change their teaching skills in teacher education courses. However, the ES teacher who regressed most towards his pre-treatment patterns of questioning in the delayed lesson was a high dogmatic (193 points pretest and 191, posttest). The in-service courses, however, did not appear to have any influence upon teachers' attitudes in terms of this scale.

Zevin (1969) reported pretreatment means of 98.3 on his Attitude-to-Inquiry scale for his experimental group of American secondary teachers (N = 15), and 90.8 for his control group (N = 10). The pre-treatment means in the present study were similar, ranging from 92.7 for Control to 99.3 for EG teachers. However, the three groups changed little on the posttests (2.9 points was the maximum — for the Control group), compared with a recorded change by Zevin of 7.3 points for his experimental group after an inquiry-based in-service course.

These scores show that the Hamilton teachers, as a group, were positively disposed towards inquiry and classroom openness, before the treatments took place. Perhaps, then, a large change could not be expected. However, these attitude scores suggest that some teachers' actual behaviour is probably at variance with their professed behaviour about how they do or should conduct inquiry lessons. The predominance of teacher talk and memory recall questions in many lessons illustrated this.

8.8 Changes in pupil attainment

Tests of pupil attainment were included in the study in an attempt to find out whether or not alterations in a teacher's questioning would result in corresponding changes in pupil learning. Numerous writers (for example, Crabtree, 1967; Taba *et al.*, 1971; Klein, 1972; McNaughton, 1974) have stressed the need for social studies teachers to devise learning experiences which cover a wide range of thinking skills, and then to construct tests to measure these. They have also expressed disappointment that, to date, evaluation in social studies has tended to be concerned with a narrow range of thinking skills, mainly recall of information. Accordingly, in the present experiment, an attainment test was constructed for each lesson package. The test was divided into four subtests to measure the thinking skills of Cognitive memory, Higher-order thinking (all levels beyond the knowledge level of Bloom's Taxonomy of Educational Objectives), Evaluative and Divergent thought. Additionally, an inference test was administered at the beginning and at the end of the experimental series.

The multiple-choice technique was used for the subtests of Cognitive memory and Higher-order thinking, and the results of analyses of pupil scores showed that the pupils of all three groups of teachers achieved at a similar level in each of the three tests (one pretest and two posttests). Although the mean score of the ES pupils for Cognitive memory dropped slightly in the immediate posttreatment test, so did those of the EG and C groups, and this could have been a reflection either of the lesson content or the particular test items used on this on this occasion.

Pilot studies of the subtests, and previous research (Elley and Reid, 1969) indicated a high correlation between multiple choice items which measure memory recall and higher-order thinking skills — such as comprehension, application and analysis — and intelligence, as measured by tests like the *Otis Tests of Mental Abilities*. It had therefore been hypothesised that any changes in teacher questioning techniques would have little effect upon pupils' Cognitive memory and Higher-order thinking scores; that is, intelligence would be the main contributor as to whether or not pupils had acquired specific items of knowledge and could utilize them inferentially. This contention was supported by the results.

The analysis of free response scores, however, revealed interesting changes. Whereas the EG and C group pupils had a similar pattern of scores for the Evaluative and Divergent subtests on the three test occasions, the ES pupils increased their scores, as is shown in Figures 7.1 and 7.2, until they were significantly different from their counterparts in the other groups.

In Chapter 3 it was pointed out that there is some evidence that teachers who encourage creativity and divergent thinking can produce almost immediate shifts in how their pupils talk (Torrance, 1962; Getzels and Jackson, 1962; Guilford, 1967), and improvements on creativity test scores can result after short-term teacher emphasis upon skills such as asking divergent questions (Franklin and Richards, 1977). It seems that the ES teachers, who, in the first posttreatment lesson, decreased their talk from an average of two-thirds to just over one-third of total talk, and who asked far more open-ended questions, somehow communicated their shift in teaching style to their pupils. The children found themselves far more actively encouraged to speak out and give their own ideas, as illustrated, in the transcript quoted earlier in the section.

When it came to answering free response items, the results show that the new encouragement to give ideas in discussion apparently carried over into the children's subsequent test performance. It appears that the children quickly changed their own behaviour in response to the teacher's change. An associated factor may have been that pupils sensed a reduction in the teacher's pre-experimental emphasis of 'getting the right answer'. The disparity between ES

pupils and those of the EG and C teachers was even more marked for Divergent category responses in the delayed test (Figure 7.2). This perhaps indicates that pupils had accepted the new patterns of interaction on a longer-term basis. It is possible, of course, that the change in pupil test responses could have been situation-specific and/or teacher-specific. It would have been interesting to have been able to analyse the same teacher-pupil variables for these classes when they had different teachers, to see whether the trends carried over.

The above comments and results can be compared with this statement by Gallagher, Aschner and Jenné (1967, p.94) about their own study:

> One of the most powerful observations made as a result of the data collected in the present study was how central and crucial the teacher's role is as initiator and determiner of the kinds of thought processes expressed in the classroom. It was the teacher who asked the directing questions which were the focus of the discussion. The way in which these questions were presented determined the kind of thought operations the students performed. The tendency of the students to respond in the same idiom was so strong that whenever it did not occur, the observer was left with the feeling that either the student did not understand the question, or something was grievously amiss in the rules governing teacher student questioning in that classroom.

Incidentally, it was noted that the free response scores of pupils did not correlate to a high degree with intelligence. This finding corresponds with those of other researchers in the field of divergent thinking. For example, Getzels and Jackson (1962) found positive but low correlations (about 0.26) between pupils' intelligence tests and divergent thinking test scores. Table 7.93 shows that in the present study, positive correlations between the scores for Divergent thinking free response items and Otis Intermediate Intelligence Test scores ranged from 0.33 to 0.43.

Finally, no significant differences were found between the ES, EG and C pupils' inference and caution scores on the *Social Studies Inference Test*, for either the pretest or posttest. This suggests that the more pervasive intellectual skills of inferential and cautious or tentative thinking are changed only over a longer time than is the

case with evaluative or divergent thinking. Also, although the teachers who changed their questioning most — the ES group — were encouraged to stress inferential thinking, the in-service course time devoted to studying this aspect of pupil behaviour was limited. Far more time was given to the concentrated study of how to ask various kinds of questions. However, inferential thinking is probably important in answering Evaluative and Divergent thinking questions, which had been stressed in the in-service programme. It would therefore be interesting to carry out further analysis of the transcripts to see whether the ES group pupils' answers contained more inferential thinking in the posttreatment lessons than in the pretreatment ones.

8.9 Summary

Evidence from the present study, then, provides support for the research hypotheses relating to teacher behaviours in that, apparently as a result of a specific skills in-service programme, the ES teachers who took that course changed certain questioning skills more than teachers who did a general in-service course on inquiry teaching in social studies. However, the EG teachers who were given the general course did not change any more than a control group who were given no treatment and the two groups remained essentially similar in their pre- and posttreatment questioning behaviour. The changes made by the ES teachers were maintained to a significant extent in lessons they taught three months after the in-service treatment. As predicted, teacher attitudes (dogmatism and inquiry) were not affected by the in-service programme.

The analysis of the pupils' test scores supported the research hypothesis that there would be no substantial differences between the scores of pupils of ES, EG, and C teachers for the pre- and posttreatment Cognitive memory and Higher-order thinking subtests. It had been predicted that on posttreatment Evaluative thinking and Divergent thinking subtests — which were made up of free response items — the ES pupils' scores would be significantly higher than both the EG and C pupils. While the ES pupils' scores were significantly higher

than the other two groups, there was no difference between the EG and C groups. These findings indicate that the teachers' greater use of Evaluative and Divergent questions in discussion of social studies topics can increase their pupils' scores for written test items which require them to use skills similar to those they would employ in answering the oral questions.

CHAPTER 9

CONCLUSIONS

9.1 Recapitulation

Through the testing of a number of hypotheses, some significant results — both statistically and educationally — have been found in this study. A number of the research predictions were realised, and the following conclusions seem warranted:

9.1.1 The teachers who were given a four-day in-service course which focussed upon specific questioning skills changed their subsequent classroom questioning behaviour to a significant extent by asking more Evaluative and Divergent questions, and fewer Cognitive memory and Convergent questions than teachers who did not participate in the course. They also talked less and changed the discussion so that there were significantly more extended patterns of pupil participation.

9.1.2 These changes in questioning and patterns of interaction were maintained for at least three months after the in-service course.

9.1.3 The teachers who were given a different four-day in-service programme, dealing with inquiry teaching in social studies in a general way, did not change their subsequent questioning behaviour to any significant extent when compared with teachers in a control group which had no in-service programme.

9.1.4 The pupils taught by the ES teachers (who changed their questioning), significantly increased their production of reasons and ideas in free-response test items compared with the performance of pupils of the general-course and control-group teachers.

These findings have considerable implications for the pre-service and in-service education of teachers. It seems that when in-service programmes set out to alter teaching skills, it is not sufficient to treat the skills in a general way by lecturing about them, nor just by giving out readings about them. Rather, in-service educators must identify the skills, operationally define them, and then clearly communicate them to the teachers so that they are understood. Then, as Borg (1975) and Taba (1966) have stressed, it is very important that teachers be given opportunities to practice the skills, be given feedback about their performance following the practice, and then be encouraged to make further modifications where necessary to improve their performance. It was with the above recommendations in mind that the SPRITE programme was planned for the present study, and the outcomes in terms of teacher and pupil changes have indicated strong justification for the approach adopted in that programme.

Furthermore, it would appear that the most effective way to teach the above skills to teachers is, as Rubin (1973) also argued, to run in-service courses in such a way that teachers come to realise the benefits for their teaching of adopting new and altered techniques. With the ES teachers in the present study it seemed that, once they had made a 'self-decision' to try to learn new ways, they then accepted help and advice, and so were in a much better position to build up a repertoire of skills to use in their classrooms — and not just a narrow range of 'tricks'.

A necessary implication of these recommended procedures — although not specifically tested in this study — is that in-service courses are much more likely to be effective if they space the attendance to give teachers a chance to practise skills in their own classrooms between in-course sessions. Failure to provide this spacing eliminates opportunity of real — as against simulated — feedback for teachers. Perhaps more extensive changes in teacher behaviour could be effected if in-service educators could also be involved in the classroom working <u>with</u> each teacher between in-service days, to help develop new skills: further research along these lines could be most productive.

The size of the in-service group appeared to be another factor that contributed to the success of the skills course in changing

questioning behaviour. Having just eight teachers enabled the investigator (who ran the course) to get to know the teachers well, and to gain their confidence. There was also more chance than in a bigger group, for every teacher to contribute to discussion, and more time for him to get feedback about his own teaching.

On the basis of the posttreatment lessons of the EG teachers, there is also some evidence from this study to suggest that merely talking about 'inquiry teaching' or 'questioning to promote inquiry', will have little or no effect on what teachers do when they return to the classroom. On most variables the teachers who took part in the 'general' course showed little change in pre- and posttreatment questioning patterns. It had been predicted that members of this group would change their questioning more than would the control teachers. It was indeed surprising to find no significant differences between them, but perhaps this reinforces the point just made concerning the importance of practice and feedback, which neither group experienced. Furthermore, the findings suggest that curriculum-makers cannot assume that changes to teachers' classroom behaviour will result from just sending to schools information about syllabus innovations and teaching techniques. The evidence from this study is that, even if such information is read, changes to classroom practice would probably not occur. Therefore, much more in-service education is needed which focusses upon specific classroom skills, as well as the broader aspects of curriculum innovations.

The question classification system devised for this study was an exploratory attempt to develop an instrument suited to analysing the inquiry methods used in social studies lessons, by focussing on various types of questions designed to get children to think in different ways. The categories included four from the Aschner-Gallagher system (in Gallagher, Aschner and Jenné, 1967), which in turn derived from the Guilford Structure-of-Intellect Model (Guilford, 1959). Two other categories — extension and grounding — were also included, together with a 'routine' category.

This system proved successful in pointing up two important kinds of teaching behaviour mentioned by Ryan (1969). First, Ryan observed that '... opportunities were lost when topics were dropped after the initial question was answered' (p.96). Second, he said that in-service

courses needed to focus on 'questioning skills which enable the teacher to facilitate children's efforts to make the "cognitive leap" from concrete factual knowledge to analysis, synthesis, and evaluation' (*ibid.*). There was a good deal of evidence, both in the posttreatment discussions and test performance of the pupils of the ES teachers, to show that the above improvements were realised, to a significant extent, when teachers used appropriately framed questions.

Two other advantages of the question-category system were evident: the categories were able to be taught easily to teachers, and coders could be effectively trained to classify the types of questions from transcripts of taperecorded lessons. The training programme showed that, in a relatively short time, a satisfactory inter-coder agreement was possible.

The development of specific lesson materials for each lesson taken by the experimental subjects (teachers), was another exploratory aspect of this study. Although the construction of television material was time-consuming, the experimental design was considerably improved; lack of standardised content in previous classroom experiments has been one of the chief sources of intervention from uncontrolled factors (Dunkin and Biddle, 1974). It could be assumed, as a result, that observed differences in both teacher and pupil behaviour were due to factors other than content.

The construction of the specially-designed attainment tests also represented a departure from earlier studies in this area. No previous research was located in which the effects of changes in teachers' questioning resulting from in-service work had been measured in terms of a range of tests of different thinking abilities in social studies. The results of this aspect of the present study, if confirmed by subsequent research, have particular implications for the classroom teaching of social studies. It appears that, when pupils have already been given information (as they were through the television presentation in this study), it is unnecessary for teachers to ask numerous Cognitive memory and Convergent questions in follow-up discussion. When teachers <u>did</u> reduce these types of question, there was no significant decrease in test scores on Cognitive memory and Convergent thinking items. It is interesting to compare this finding with that of Wright and Nuthall (1970) who found a positive relationship between the proportion of closed questions asked by teachers and pupil attainment in science lessons.

It was a particularly important finding of this study that the increased asking of Evaluative and Divergent questions by teachers significantly increased pupils' ability to handle the free-response items in the attainment tests. A surprising aspect was the immediacy of the changes observed, with pupils contributing more ideas both in discussion and in written responses in the tests. The findings suggest that pupils are quick to pick up 'cues' from the teacher to alter their own behaviour. Nuthall (1970, p.28) had recognised pupils' ability to do this: 'pupils can and do make use of subtle cues from the teacher'. Thus, from the teacher's point of view, it seems that pupils' creativity and curiosity can be fostered by asking open-ended questions and giving pupils more time to think, and that this transformation can take place quite quickly.

Finally, the 'second-step' benefits which resulted from the SPRITE programme represent a challenge for in-service educators, because they show quite clearly that an in-service course can change teachers' specific verbal behaviours, and these altered behaviours can in turn improve pupil learning — first, by producing better discussion, and second, better performance on tests requiring complex thinking. It would indeed be quite possible to run a programme such as the SPRITE course entirely in the school setting; that is, as a school-based 'package'. The programme could be run either by outside educators who would come into a school and work with several staff at once, or at courses away from the school. Teachers could be trained to become 'experts' in the techniques, and they could then go back to their schools to conduct programmes with other teachers. However, it would seem important that teachers should be released from their classes, so that extended periods (such as four or five days) could be spent working on the skills, interspersed with periods in the classroom for practice. It is also possible that several schools could send teachers to one school in close proximity for such a course. However, for these kinds of programmes to succeed, there would need to be provision of adequate resource materials, such as taperecorders, lesson materials, and perhaps videotape equipment.

The above findings, then, relate to a number of theoretical issues which were outlined in Chapters 2, 3 and 4. Analysis of base-line data indicated that — like teachers in other studies — the New

Zealand teachers who made up the sample for the present study tended to dominate the talk in their classrooms. In the social studies lessons which were taperecorded the teachers asked mainly questions which required their pupils to recall information and few questions which called for the use of higher-order thinking in their pupils. Researchers (for example, Massialas and Zevin, 1967) have noted that this probably limits the degree to which inquiry learning occurs in class discussion. Other researchers (for example, Borg, 1975) have shown that through in-service techniques teachers' questioning can be altered, and this was also demonstrated in the present study. As a result, the ES teachers engaged their pupils in responding to a much wider range of questions. It can be argued that their classes were participating in an inquiry-based discussion to a far greater extent than they had been previously. It seems too, that these changes are able to be maintained over at least a three-month period.

9.2 Limitations of the study

At this point it is necessary to remind oneself of the limitations inherent in the study. It had to be restricted to a relatively small sample of Intermediate school teachers and pupils. The teachers were all males, and the three experimental groups had to be matched to ensure a basic initial comparability. However, subsequent multiple analysis of variance and one-way analysis of variance of initial observations showed that there were indeed no significant differences between groups for baseline, pretreatment data, there being a similar degree of variance within each, as well as between them. In addition, because 80% of the male Intermediate school teachers volunteered to take part in the study, the twenty-four teachers finally selected could be considered reasonably representative of that population. Moreover, since the pupil sample was drawn only from the class levels of Form One and Two, the results of the study should not be generalized beyond the age range of eleven- to thirteen-year-olds.

Some of the measuring instruments devised need to be subjected to further experimentation before too much can be claimed for them. For instance, the questioning system proved to be suitable for recording and analysing the range of questions employed in inquiry lessons in

social studies, but although the coders of transcripts reached a satisfactory level of agreement for each category, there was still an error factor, due to the necessity to make inferences about both the various thinking skills involved in pupils' answers, and teachers' intentions when asking a particular question. The patterns of classroom questioning recorded in this study arose out of specific lesson materials for cross-cultural topics in social studies. Therefore, the behaviours found may not be generalizable to all other subjects or even to all aspects of social studies.

Furthermore, it is difficult to establish reliability, validity and standardised marking procedures for free-response items like the ones used in this experiment; although a high degree of consistency was achieved in the marking of the tests in this study some caution must be exercised when interpreting these particular results. And, finally, the variables analysed in this study represent but a small part of the total number of variables (teacher and pupil) which make up the very complex interaction of the classroom.

9.3 Directions for further research

Obviously more research is needed before the findings of this study can be generalized at all widely. For instance, Ryans (1960) observed differences in teaching behaviour between male and female teachers, and in some cases between older and younger teachers. It cannot be assumed, without further experimentation, that the in-service programme used in this experiment for the Experimental Skills group would be equally successful in all circumstances.

This experiment has produced interesting findings about a particular social studies teaching-learning situation, but there would appear to be considerable scope for further research into inquiry methods in the social studies and in other subjects. For example, would teacher questioning behaviour be affected by setting 'advance organizer' questions for the class before the teacher presented his content material? What effects would this have on pupil motivation, and on the subsequent discussion? And pupil questioning, as well as teacher questioning needs to be investigated to find out whether children can become more effective and fluent at

seeking information from the teacher and from each other if actively encouraged to do so.

Further research is also required to evaluate other kinds of in-service programme, and to measure the impact they have on teachers' classroom skills. Varying durations of courses should be tried, and different spacing of times of attendance; for example, how long do courses have to be to change teaching skills such as questioning? How many follow-up contacts are needed between in-service educator and teacher to reinforce these skills? How long do teachers maintain the skill changes induced by different kinds of in-service programme?

In addition, the testing of pupils achievement in social studies represents a challenging area for further research. Some important objectives can be measured only by free-response items, and so more free-response tests are needed to explore the thinking abilities of children of various ages. However, such tests are difficult and time-consuming to construct, and team research would seem more appropriate than one-researcher studies, so that ideas can be shared, and a more comprehensive experimental programme be carried out.

9.4 Summing up

In the present study an attempt was made to find out whether two kinds of in-service courses would modify teachers' questioning behaviour in social studies lessons, and whether any such changes would in turn affect pupil attainment. It was found that teachers who were given a skills-based programme were able to significantly reduce their asking of Cognitive memory and Convergent questions and to increase their use of Evaluative and Divergent kinds. In class discussion they also talked much less than they had prior to the treatment, and encouraged more extended patterns of pupil involvement. These changes were maintained to a significant extent in lessons taken three months after the in-service course. Teachers who took a general-type in-service programme did not change any more than teachers in a control group. The pupils of the specific-skills group teachers were able to significantly increase their production of answers to free response items measuring evaluative and divergent thinking, without decreasing their scores on cognitive memory and higher-order multiple-

choice tests.

The results of this experiment demonstrate that particular teaching skills, such as questioning, can be successfully taught in in-service programmes by a procedure of sensitizing teachers to the skills, practising them, and reviewing competency through feedback techniques. When questioning and associated skills are taught in this way, improvements in the range of pupil learning also results.