AN INVESTIGATION INTO THE RELATIONSHIP BETWEEN STUDENT CHARACTERISTICS AND THE INQUIRY METHOD OF TEACHING SCIENCE

by

Udo Westerneng, B.Ed., B.App.Sc. (Mar. Eng.), Dip. Eng. (Mar. Eng.)

A thesis submitted in fulfilment of the requirements
for the degree of Master of Education (by Research)

School of Education, Launceston
University of Tasmania

September 1993

TABLE OF CONTENTS

LIST OF DIAGRAMS

LIST OF TABLES

ACKNOWLEDGEMENTS

On completion of this research, I would like to sincerely thank the following people:

Mrs. Noelene Westerneng, for tolerating the loss of weekends and holidays when this research took priority over all other activities and for her constructive ideas; Dr. David Hannan for his tireless reading of numerous drafts and for his detailed comments.

DECLARATION

I certify that this dissertation contains no material which has been accepted for the award of any other degree or diploma in any institute, college or university, and that, to the best of my knowledge and belief, it contains no material previously published or written by another person, except where due reference is made in the text of the dissertation.

Udo Westerneng

PERMISSION TO COPY

I hereby give permission to the staff and students of the School of Education within the University of Tasmania to copy this dissertation in whole or in part without reference to me. This permission covers only single copies made for study purposes, subject to the normal conditions of acknowledgement.

Udo Westerneng

ABSTRACT

Five instruments were used to collect data on student characteristics: the Combined Cognitive Preference Inventory, a modified Cognitive Style of Categorization Behaviour, the Test Of Science Related Attitudes, the Structures of Observed Learning Outcomes and a designed Student Perceived Characteristics of Success instruments.

In this research five classes were taught science using the inquiry method. One of the treatment classes was matched to a control class taught conventionally. The data collected were analyzed in two different ways. First, the data of the five treatment classes were analyzed using non-parametric statistics. It was found that students who coped well with the inquiry method had significantly different characteristics than students who did not cope well. The distinctions between successful and unsuccessful inquiry students differed significantly between ability levels. A model was developed to show the hierarchal nature of inquiry student characteristics at different stages of cognitive development.

Second, the data of the matched control and treatment classes were analyzed using parametric statistics. Student characteristics were incorporated in linear regression equations and it was found that the regression equations for control students were significantly different from the regression equations for inquiry students. Also, the inquiry method produced significantly better end of quarter achievement scores than the control method. This was not true for all achievement criteria at S.O.L.O. Singular and Multiple levels. The model developed from the non-parametric analysis was tested and refined by the results of the parametric analysis.

CHAPTER 1

INTRODUCTION

Curiosity and intriguing observations have spurred on scientific inquiry and the search for scientific knowledge in the past and will do so in the future. Curiosity and the tendency for the search for knowledge to start with free exploration always features in the Learning Cycles propounded in education circles (Bassler and Kolb, 1971, 112; Soloman, 1980, 33; Driver, 1983, 51; Fisher, 1985).

If these learning cycles reflect both the natural and the desirable state for learning, why use science teaching methods like the expository and the deductive methods which allow no exploration and disequilibration? If inquiry is both natural and desirable (Welch et al., 1981, 37; Tamir, 1985, 93) why is this method unsuccessful in achieving its aims (Chapter 2, page 37) and why has inquiry been used very little in science curricula (Shulman and Tamir, 1973, 1112; Welch et al., 1981, 37)?

Although it is partly known why students find the inquiry method difficult (Igelsrud and Leonard, 1988, 305) it is not known why some students are better able to cope with the inquiry method than other students (Welch et al., 1981; Seymour et al., 1974; Rotheram, 1984; Schwab, 1966). If the characteristics of successful students were known these characteristics could then be developed and promoted to give more students a chance of success.

This research into the inquiry method of teaching science was centred around two focii:

(1) could the characteristics of successful and unsuccessful inquiry students be precisely formulated in an equation?

(2) were the characteristics of students in inquiry-based courses different from the characteristics applicable to students enrolled in conventional science courses? Furthermore, for easy visualisation, could a model, depicting student characteristics required in inquiry-based courses, be developed from the answers to this question? The latter question was not the main question because such a model would depend on the answers to the earlier question.

The formulation of an equation does not necessarily imply that students are programmed according to this equation. An equation simply allows the relevant characteristics to be separated from irrelevant ones. Ideally, an equation should also have predictive powers, but the identification of relevant characteristics is the main focus of this question. To facilitate communicating the identification of the relevant characteristics a model could be developed. The visual nature of a flowchart like model would overcome the mathematical nature of an equation. Unfortunately, a model does not show the relative importance of each of the characteristics as clearly as an equation. These focii became the basis of this thesis and shaped the design of this research.

Research focus (1) was split up into three separate, but related, issues:

extracting the relevant characteristics of successful and unsuccessful students, investigating the relevant characteristics with regard to student ability and quantifying the relationship between relevant characteristics and science subject matter, namely chemistry, physics and biology.

Research focus (2) was divided into two parts: comparing inquiry to conventional science teaching methods on the basis of achievement scores and formulating equations which show the relative importance of each of the characteristics found in the first research focus. The data developed in the first research focus could be used to construct a model which showed visually the relationship between the relevant characteristics. The findings of the second research focus could then be used to check and refine this model.

The body of students needed to investigate both research focii required a large number of students taught by the inquiry method of teaching science. A large number would be needed in order to incorporate: successful and unsuccessful inquiry students, students of different ability and the three main stream science subjects (Chemistry, Physics and Biology). Another requirement of the research was that the inquiry method of teaching science had to be compared to a conventional method of teaching science. To achieve this six classes were used: five treatment (inquiry) classes and one control (conventional) class. One of the five treatment classes was matched, with respect to student age, ability and subject matter, to the control class.

Statistically, two approaches were followed. The first approach used only the five treatment classes in a non-parametric analysis to investigate research focus (2). The second approach used the matched treatment and control classes in a parametric analysis to investigate research focus (1).

Although this research limits itself to investigating only one method of teaching science, it will provide precise information on which and in which order student characteristics should be developed to give more students the opportunity of being successful when taught by the inquiry method of teaching science. Eventhough this method is only one in a range of teaching methods, it is generally not taught well (Welch, 1981, 37; Igelsrud and Leonard, 1988, 305) and the findings of this research could change that situation.

The implications for classroom practices, based on the results of this study, are that students of different ability levels must be taught differently when using the inquiry method of teaching science. They must be taught differently by focussing on different student characteristics. For instance, students at the Piagetian concrete stage of development characteristically benefit if the object or phenomenon being studied is presented as a whole (as opposed to presented as a combination of interrelated parts) and investigated with regard to application and function.

In overview, this thesis is presented in six chapters. In this first chapter the problem setting is presented in minor detail, the research problem is defined and the rationale, both conceptually and methodologically, is outlined. Chapter two looks at what inquiry is and what research has found out about the methods of teaching science, including the inquiry method. Chapter three describes five instruments which were used to collect data on the students who participated in this research. Two of the five instruments were developed as part of this research. Chapter four explains and justifies the design of this research. The level of detail is high in order to facilitate replication for future research. Chapter five outlines in detail the parametric and non-parametric statistics used to analyse the data of this research. Chapter six presents the conclusions that were drawn from the results of the statistical analyses. From the non-parametric statistics a model is developed showing the hierarchal nature of the characteristics of inquiry students. Linear regression equations are developed from the parametric statistics. These regression equations are also used to test the predictive power of the model developed in the non-parametric analysis of the research data and provide a means for refinement of the model.

CHAPTER 2

LITERATURE REVIEW

2.1 Outline of chapter two

This chapter is divided into three parts. The headings of these parts are printed in bold text for clarity:

(1) explanation of terms used in the literature;
(2) relationship between science and science teaching;
(3) the methods of science teaching.

In part three, each of the methods of science teaching is discussed in turn. First, the aim and definition of each method are stated. Second, the results of research studies are presented. The results of the research studies are given in three parts:

(1) the advantages of the method;

(2) the disadvantages of the method;

(3) any ambiguous conclusions that were reached.

2.2 Explanations of terms used in the literature

In this section some of the relevant terminology used in the literature will be clarified.

2.2.1 Inductive versus deductive science

In mathematics it is illegal to proceed to the next line if that next line does not follow on logically or axiomatically from the previous one. This is called deductive reasoning. In science, at least in experimental science, most of the time explanations are proposed after data have been collected. This reasoning from data is called inductive reasoning. Although there are many philosophers and scientists debating which of these reasoning styles is better (Sakmyser, 1974, 67; Hermann and Hincksman, 1978, 37; Quinn, 1983, 38; Sternberg, 1986, 281; Davson-Galle, 1989), at high school students taught to reason inductively have produced significantly better results in critical thinking than students taught to reason deductively (Bates, 1978, 62).

2.2.2 Process and method

By process is meant the way in which scientific information, also referred to as the body of knowledge related to science, is obtained. This is quite different from the actual information itself. Scientific methodology, that is experimental methodology, is another term often coined in connection to process. Methodology refers to the axiom of controlling all variables except one. This axiom is implied by the term experimenting whereas investigating, because it is a more general term, could imply no more than information gathering. Although the term scientific method is often applied there is, in fact, not one such method but a variety of them; the section: *the choices a scientist faces*, of this literature review chapter will develop this further.

Regardless of the method used the following three words are always used: hypothesis, procedure, conclusion. An hypothesis is a definitive statement that describes a proposed relationship between variables. The hypothesis can be derived at by pragmatic, theoretical or practical application considerations, but the generation of intuitive ideas has been suggested as important (Wilson, 1974, 131). Wilson (1974, 131) placed a lot of emphasis on 'hunch generation' and used this to dismiss the idea of 'students acting as little scientists' because they lack the conceptual background. Wilson believed it acceptable to let students act as semi-scientists being guided into useful directions by the teacher. Wilson's preferred technique is to make the students conversant with the skills involved with the processes of inquiry, noting that effective inquiry will always be '...limited by the prerequisite knowledge of the learners...(and that) teachers need to determine previous knowledge and select appropriate stimuli' (Wilson, 1974, 132). A procedure is a series of steps that are to be followed sequentially as the investigation proceeds. A conclusion is the interpretation that is made to explain the results that have been collected in the experiment.

2.2.3 Fluid and stable inquiry

To gain further understanding of what science as an inquiry is about Duschl (1986, 28) distinguished between fluid and stable inquiry. The latter refers to refinement of existing knowledge whereas the former describes the development of new concepts and theories. If students are to understand science as inquiry both these types should be reflected in the science curricula. Because fluid inquiry is the type that provides the breakthroughs in science, Schwab (Duschl, 1986, 30) argued that emphasis should be on fluid inquiry especially since by the time high school students might actually have to use their knowledge it is likely to be obsolete.

2.2.4 The choices a scientist faces

A scientist hardly starts an investigation by deciding out of hand to do some fluid inquiry. In their pursuit of knowledge scientists have to make many choices about the form the research will take. Westbury and Wilkof (1978, 184-228) edited an analysis done by Schwab on what scientists do. The first choice faced by scientists is with regard to principle of inquiry; breaking the whole down into parts, combining the parts to find the whole principle, comparing the whole principle to other principles, proving the principle does not really exist or looking for exceptions. Building a type of model is the second choice: construct a web of concepts, extract if-then relationships, illicit analogies and comparisons. If a principle has been found and a model has been built, the research is judged on four criteria:

(1) its interconnectivity to existing knowledge;
(2) its adequacy;
(3) its feasibility;
(4) can old principles be explained by the new principle, do the old principles need to be re-written and can new knowledge arise from the new principle?

The latter three are labelled continuity.

Reliabilty and validity considerations, including repeatability and precision, will determine the success of the research. Only if the research is successful, that is when the scientist approaches the concluding stage, does the scientist consciously decide on either fluid- or stable inquiry. The scientist then has to ask: am I trying to draw conclusions from an hypothesis, or stable inquiry: am I testing and refining a conclusion? Having made this decision the scientist has a choice of three tasks:
(a) 'guiding, collecting, interpreting' (Westbury and Wilkof, 1978, 218). This

refers to an invention and proposal of principles, the accumulation of data and the reviewing of data with the aim of interpreting, respectively;
(b) 'postscript on a grand strategy' (Westbury and Wilkof, 1978, 220). In this task the researcher redefines or modifies existing principles;
(c) 'speculation on a grand strategy' (Westbury and Wilkof, 1978, 222). Here principles are attacked in order to destroy them.

A less constructive type of research, designated point 0 by Westbury and Wilkof (1978, 224), is the attack on the validity, reliability, feasibility and status of a principle. This is more an attack on the researcher than attack on the principle and should not be considered as an option.

Wilson (1974, 128) did not use the terms fluid and stable in his analysis of how scientific inquiry occurs. Instead, Wilson used the ideas of empirical and conceptual inquiry. The former is a search for quantifiable information, the latter is a search for related variables such as properties and attributes. Wilson (ibid) explained that both are initiated by stimuli that lead to search processes and "finish" in results. Empirical inquiry is stimulated by discrepant events such as a curious happening, data gaps, a chance observation. The stimulus for conceptual inquiry is a discrepant attribute, such as a contradictory phenomenon, the search for a limit, the proposal of a theory. The key to linking Wilson's model to Schwab's categorisation would appear to be the stimulus for Wilson's inquiry: a discrepancy. Stable inquiry takes something that is accepted as knowledge and refines it, may even redefine it without destroying the whole. A discrepancy implies a lack of knowledge, therefore, Wilson's model is concerned with fluid inquiry, with, perhaps, conceptual inquiry being closer to stable inquiry than empirical inquiry. The result of empirical inquiry is an exposure of a new phenomenon, the result of conceptual inquiry is a new explanation. Note that both are new, again suggesting fluid inquiry.

2.3 The relationship between science and science teaching

2.3.1 Skills needed in science

Science can be taught in many different ways. The pathway to achieving the goals and objectives set for or by the teacher is never a prescribed one. A multitude of barriers, limitations, restrictions and available skills, equipment and resources may seem to force the teacher into returning to tried and proven methods. However, a skilled and aware teacher has many alternatives available. The different methods fall into broad categories such as: expository, deductive, inductive, demonstration, laboratory, discovery, problem solving and inquiry. Some of these categories can be further subdivided on a continuous scale with one extreme closed-ended and the other open-ended.

Each of the methods listed have their own emphasis on what they regard as important for student needs. The expository method regards knowledge as being most important, whilst the deductive and inductive methods regard reasoning as the most important.

Regardless of the emphasis the method places on teaching, the students are acquainted with a number of skills which a scientist uses. A summary of these skills was constructed by a committee of the Tasmanian Schools Board (Pallett, 1983, 2).

These skills can be generalized into seven categories:

(1) observation;
(2) data collection and treatment;
(3) classification;
(4) interpretation;
(5) model construction;
(6) problem solving;
(7) identifying assumptions.

Because these skills are deemed essential to scientific work (Pallett, 1983, 2) it seems clear that no matter which method is used to teach science the students must become familiar with as many of these skills as possible. Therefore, the relationship between science and science teaching is a complex one in which the aims and methods of science do not always run parallel to the aims and methods of science teaching. This complexity is highlighted in the tension between knowledge and process. Some questions which arise from this tension are: should students be like little scientists? Is factual knowledge more important than trying to act like scientists? The answers to these and other questions will, in part, determine which teaching method is to be employed in the teaching of science.

2.3.2 Open- versus closed-ended investigations

The terms open and closed refer to whether the objective of the investigation is stated in terms of precisely achievable objectives, such as: the students will be able to separate salt from water by means of evaporation, or whether the objective leaves room for variation in the means to achieve the objective, such as: the students are to separate salt from water by whatever means they consider suitable.

If open- and closed-ended are viewed as a dichotomy, then four methods are possible (adapted from Shulman and Tamir, 1973, 1112).

Table 2.1 An array of possible closed- and open-ended laboratory exercises.

Method	Hypothesis	Procedure	Conclusion
closed	closed	closed	closed
open conclusion	closed	closed	open
closed hypothesis	closed	open	open
open	open	open	open

In table 2.1 the concepts hypothesis, procedure and conclusion are used. These concepts were defined at the start of this chapter under the heading: *2.2.2 process and method* (see page 6).

It can be seen in table 2.1 that if the hypothesis is open the other two must necessarily be open as well. If a particular method of obtaining data (procedure) is stipulated the hypothesis is consequently restricted (closed). An example of an open hypothesis is: what are the variables that effect the behaviour of a vapour? This might, for instance, be an appropriate start to an investigation into vapours after the students have finished studying the behaviour of gases. Clearly, the teacher cannot accurately predict how the students are going to tackle the problem. Superheated steam is a gas, saturated steam is a vapour. Students could focus on P-T or V-T relationships of vapours, or they could focus on problems with the use of vapours. That is the procedure is open, nor can the teacher predict the outcome of the investigation, that is the conclusion is open as well.

It is apparent that the relationship between science and science teaching involves a complex web of interacting relationships. Doing science and learning about science have similarities but also large differences.

2.4 The methods of science teaching

This thesis evaluates the open-ended inquiry method of teaching science. The following sections describe the perspective of the range of science teaching methods. The first part introduces brief definitions of the laboratory, expository, discovery and inquiry methods. The reason for selecting the first three methods is that they are the main categories of science teaching methods used in high schools (Tinnesand and Chan, 1987, 43; Mulopo and Fowler, 1987, 218; Renner, 1976, 222; Gallagher, 1987, 352). Current teaching methods are readily classified as at least approaching one of these methods.

The laboratory-based approach is analysed in some detail because the laboratory was mainly used to teach open-ended inquiry science in the cause of this research. Although the laboratory was the medium in which the inquiry method was taught, the method of teaching was exclusively by inquiry. This discussion of the laboratory method is also important as large amounts of money are expended on laboratories (Sund and Trowbridge, 1967, 91) on the unquestioned assumption that the laboratory is the most appropriate medium by which to teach science. The ambiguous conclusions from research into teaching by the laboratory method are presented.

The last section relates to the inquiry method and, where appropriate, outlines research outcomes that have resulted from studies comparing it with other methods. This method is quite separate from the other three methods not only because it is used very little but because it is the method that is the focus in this research at the exclusion of any other method.

2.4.1 Brief definitions of four science teaching methods

Although there are many definitions of the methods mentioned above at this stage the following is offered as a starting point.

The laboratory method: 'a method of instruction in which the teacher presents the concepts, as well as the procedural instructions for the verification, in a printed laboratory manual, and provides each student all of the equipment for verifying each of the individual concepts' (Babikian, 1971, 201). Tamir and Lunetta (1981, 477) also use the words: illustration, demonstration and verification, in their definition and show that inquiry and problem solving are related by different teaching techniques.

The expository method: by stating the concept, illustrating and rehearsing with examples, the students are told how to interpret concepts (Babikian, 1971, 201). Often this style of teaching is presented as the method which is the most teacher directed. Other methods are frequently pitched against this method in comparisons how students perform in cognitive and affective aspects (Roadrangka and Yeany, 1982, 2; Sugrue and Thomas, 1989).

The discovery method: in its purist form the discovery method merely presents the student with materials. The student are expected to manipulate the materials until they are acquainted with them. They may then start to control some of the variables. The method is intended, also, to familiarize students with how scientific information, consisting of knowledge and understanding, is processed (Burns and Ellis, 1970, 105).

The inquiry method: 'a set of activities directed towards solving an open number of related problems in which the student has as his principal focus a productive enterprise leading to increased understanding and application' (Seymour et al., 1974, 349). This relationship between inquiry and problem solving is also put forward by Pugliese (1973, 24) and Humphreys (1978, 435). Althuogh the use of inquiry skills has been advocated (Lumpe and Oliver, 1991, 345) expectations are seldom fulfilled (Welch et al.,1981,33).

2.4.2 The laboratory method

People, especially students, associate science and science teaching with working in a laboratory. Many arguments (Page, 1983, 372; Dishner and Boothby, 1986, 49; Nielsen, 1986; Creek and Vollmer, 1991; Foshay, 1991, 3) have been presented to reinforce this association. This has not always been the case. Controlled experimentation using real objects is something relatively new to science. In ancient Greece science was conducted in the mind only (Shulman and Tamir, 1973, 1113). The use of the laboratory in science teaching is not easily justified on the basis of students' results in paper-and-pencil tests when compared to other teaching methods except that the variance of test results was greater among students who were exposed to the laboratory (Bates, 1978, 58), that is the results displayed a greater variety of responses than other science teaching methods. Furhtermore, out of 358 U.S. science teachers only 41% felt that their students learned a lot in the laboratory (Hounshell, 1985, 2). Nevertheless, the value of the laboratory with regard to the attainment of physical skills has never been disproved (Richardson and Renner, 1970, 78; Igelsrud and Leonard,1988, 303).

Aim and definition of the laboratory method

During the 1960's, efforts were made to expose students to realistic problems which could be solved in the laboratory. It was hypothesized that students' skills in problem solving, critical thinking and understanding the nature of science would be enhanced. Evaluation studies of this hypothesis have been ambiguous (Bates, 1978, 57).

Shulman and Tamir (Shulman and Tamir, 1973, 1119; Bates, 1978, 57), in the 1960's, categorized laboratory objectives into five broad areas:

(1) skills: manipulative, inquiry, investigative, organizational, communicative;
(2) concepts: hypothesis, theoretical model, taxonomic category;
(3) cognitive abilities: critical thinking, problem solving, application, analysis, synthesis, evaluation, decision making, creativity;
(4) understanding the nature of science: the scientific enterprise, scientists and how they work, multiplicity of scientific methods, interrelationship between science and technology and among the various disciplines of science;
(5) attitudes: curiosity, interest, risk taking, objectivity, precision, confidence, perseverence, satisfaction, responsibility, consensus and collaboration, liking science.

Shulman and Tamir (1973, 1120) noted that these objectives are not specific to the laboratory but also apply to teaching science in general and evidence to prove that the laboratory method alone achieves these objectives is often ambiguous.

Babikian (1971, 201) defined the laboratory method as: 'a method of instruction in which the teacher presents the concepts, as well as the procedural instructions for the verification in a printed laboratory manual, and provides each student all of the equipment for verifying each of the individual concepts.' Coulter (1966, 185) called this definition the deductive laboratory method. Others labelled this definition the verification, or cookbook, laboratory method (Bates, 1978, 56; Janners, 1988, 32). Whether or not the laboratory method is used in the closed-ended fashion, like in the definition, or in an open-ended fashion (Ramsey and Howe, 1969, 76), in general the concept or hypothesis under consideration is worked out or specified first (Babikian, 1971, 201). A procedure is agreed upon or prescribed next, a practical investigation follows and is rounded off with a conclusion which is related directly to the concept or hypothesis. This general procedure, in practice, can also be varied by working as individuals or by group demonstrations. Other possibilities include dividing the one concept into a series of related investigations or verifications. In the latter case, groups or individuals can work through the series consecutively. Alternatively, the class may be divided and the students report their investigation to the class.

Results of research studies

Studies which showed advantages in the laboratory method

Ramsey and Howe (1969, 64) and Bates (1978, 59) stated that the laboratory method was an effective method of transmitting knowledge and was found to enhance the development of critical thinking. Laboratory method experiments can either acquaint students beforehand with the correct results to be obtained (deductive verification) or students are not beforehand told the correct results (induvtive). Marie (cited in Ramsey and Howe, 1969,64), in a ten year study, found that the inductive laboratory method was superior to the deductive laboratory method on a measure of attitudes. This finding was related only to students with low interest levels. Bates(ibid), in his study, extended this finding to all students regardless of interest level.

Hofstein and Lunetta (1982, 201) reported the only significant difference between laboratory, demonstration and discussion methods was observed in laboratory manipulative skills. Students of the laboratory method were significantly better on these skills. Sherman and Pella (1969, 303) compared hands-off to hands-on physics courses; the non-manipulative groups were only shown slides of the laboratory activities. A Laboratory Skill Test showed the manipulative group was superior on manipulative skills.

Yager, Engen and Snider (1969, 85) reported that the laboratory method developed more skills with laboratory materials and procedures. Friedler and Tamir (1986, 263), in an attempt to rationalize why the laboratory method failed to live up to expectations, argued that the laboratory method's real aims were to gain 'the development of basic process skills, a "feel" for natural phenomena by adequate experiences, and problem-solving ability by adequately planned investigations....(and) emphasize the importance of discussion and reflection in making learning in the laboratory more meaningful' (Friedler and Tamir, 1986, 263). These "real" aims were included in Shulman and Tamir's categorization of laboratory aims (1973, 1112 and cited in Bates, 1978, 57) and it seems that the only difference was the emphasis on discussion. Again, the general laboratory method procedure was adhered to but much more emphasis is placed on pre- and post-laboratory discussions. Indeed, prior to any experimentation Friedler and Tamir (1986, 265) suggested non-manipulative book-work like exercises to help the students familiarize themselves with concepts on which the experiments are based. While this seems a departure from the true laboratory method, the reason for including these book-work exercises was to ensure an understanding of the laboratory method's first step in the general procedure: the introductiuon of the concept or hypothesis. Friedler and Tamir (1986, 267) claimed a significant improvement in inquiry skills as determined by pre- and post-tests, but the question that remains unanswered is: is this (bastardized) laboratory method significantly better than the other methods, such as expository, inquiry and discovery methods?

Studies which show disadvantages in the laboratory method

Hall and McCurdy (1990, 625) noted that the lecture-demonstration method was especially effective among low achievers and that the laboratory method produced a greater variance of results in end of treatment tests.

Friedler and Tamir (1986, 264) constructed a list of five shortcomings of the laboratory method:

> (1) only rarely are students required to identify and formulate problems, hypotheses, design experiments and work according to their own design;
> (2) the perceptions of students and teachers are different with regard to the goals in the laboratory and students ranked inquiry skills quite low in importance;
> (3) serious deficiencies regarding conceptualisation and understanding;
> (4) students have substantial difficulties in planning investigations, namely, in formulating questions and hypotheses as well as in designing controlled experiments.

Studies which show ambiguous conclusions

Richardson and Renner (1970, 78) reported that no significant differences were found in end of course results on the basis of exam results in chemistry. However, the value of the laboratory on the basis of laboratory skills was proven. Mills (1981, 365) in a report of an open-ended laboratory method presented an argument in favour of this form of investigation. The general laboratory method was adhered to.

Janners (1988, 32) argued that the scientific process is best taught in the laboratory. As was the case with laboratory concepts, there is no unambiguous evidence in the literature that process can be taught best in the laboratory or if it can be taught at all (Bates, 1978, 55; Ramsey and Howe, 1969, 63).

Perhaps one reason why evidence surrounding the laboratory method is ambiguous is explained by research conducted by Robinson and Tamir and Glassman (1971, 314). In both projects paper-and-pencil tests were specifically designed as a laboratory practical examination. In both projects only low (0.33) to moderate correlations were found between marks for the designed laboratory tests and conventional end of semester paper-and-pencil tests. Apparently, the end of semester test scores had little relevance to the laboratory work. Hence, the ambiguity may arise from the possibility that the different instruments used in different research projects measure different aspects of student competence in the course. To further complicate matters, Tamir and Glassman (ibid and cited in Bates, 1978, 65) noted that 'poor manipulative skills did not necessarily imply low investigative skills'. To evaluate the degree of achievement in process skills would require a prohibitive number of different instruments, especially when the interactions amongst the tests with regard to reliability are taken into consideration.

Although Ramsey and Howe (1969, 64) did not expressly explain the difference between the problem-solving and laboratory methods, they stated that the laboratory method was not significantly different from other methods with regard to effectiveness. In other student outcomes (understanding the scientific enterprise, critical thinking), no conclusive differences between the laboratory and other methods were found.

Hofstein and Lunetta (1982, 201) also reported no significant differences between laboratory, demonstration and discussion methods in student outcomes in achievement, attitude, critical thinking and knowledge of the process of science as measured by standard paper-and-pencil tests.

Hofstein and Lunetta (1982, 205) and Ramsey and Howe (1969, 68) all reported that the teacher variable was a factor more important than the external instructional mode that was required to be taught. Deductive-oriented teachers taught investigatively regardless of the form of the activity taking place.

Sherman and Pella (1969, 313) compared hands-off to hands-on physics courses. The non-manipulative groups were shown slides of the laboratory activities. No significant differences were found for measures of Watson-Glazer critical thinking, understanding science, knowledge, interest in science. It is not clear if Sherman and Pella expected differences in critical thinking or understanding. However, Welch et al. (1981, 38) observed that 'when hands-on experiences are provided, they are not characterized by true problem-solving', problem-solving and laboratory centred instruction being assumed to enhance the development of critical thinking especially (Ramsey and Howe, 1969, 66). Coulter (1966, 186) found no results in favour of the deductive laboratory approach, but neither inductive nor inductive-demonstration were superior to deductive.

Hall and McCurdy (1990, 625) noted that when the lecture-demonstration method was compared with the laboratory method, overall, there were no significant differences on a measure of the Reed General Science Test. Sherman (1989, 55) found no difference on content achievement when an individualized laboratory group was compared to a lecture-demonstration group. Bybee (1970,160), in a similar study, also found no difference in content achievement. Yager, Engen and Snider (1969, 85) found no differences in Watson-Glazer critical thinking, understanding science, attitude or knowledge, when they compared the laboratory and lecture-demonstration methods.

Hofstein and Lunetta (1982, 204) pointed out the doubtful validity of some of the standardized tests used. For example, the Watson-Glazer Critical Thinking Appraisal is not designed for general science teaching and is totally unrelated to laboratory work (Hofstein and Lunetta, 1982, 204).

The shortcomings of the laboratory method, as reported by Friedler and Tamir (1986, 264), could be inadequacies of the teachers' teaching skills and reinforced the overriding importance of the teacher variable. Students experiencing difficulties with planning investigations and controlling experiments is more a student development related factor, especially with regard to controlling variables. This suggestion is confirmed by a study by Egelston (1973, 476 and also Zingaro and Collette cited in Bates, 1978, 62) in which no significant differences were found in subject matter learned, general critical thinking (Watson-Glazer) and understanding of science between two groups, one taught by the inductive laboratory method, the other deductively. Only a significant difference in an investigator-designed measure of critical thinking was found in favour of the inductively taught group. Zingaro and Collette (cited in Bates, 1978, 62) and Shymansky and Penick (1981, 412) concluded that some teachers worked more effectively with one method than another.

Concluding remarks on the laboratory method

The laboratory method has been discussed in great detail with several key definitions given. It has been demonstrated that there is no such thing as one true laboratory method, but rather a number of methods share common characteristics.

The 1978 state of the use of the laboratory in U.S. high schools was investigated by Tamir and Lunetta (1981, 477). They reported that '...many teachers prefer to run "smooth" demonstration and verification type laboratories'; no doubt these teachers feared discipline and management problems. Costenson and Lawson (1986, 151) found ten reasons why inquiry is not taught more often: too much teacher time and energy, too slow, too difficult, risk too high (student failure), no formal thinkers, students too immature, teacher teaching habit, inquiry too sequential, teacher and students feel uncomfortable, too expensive. None of these ten reasons need to be true, but teachers believe these are the failings of the inquiry method.

Tamir and Lunetta's findings (1981, 482) are summarised below in three main points.

> The laboratory investigations that were examined:
>
> (1) almost all investigations are highly structured;
> (2) seldom, if ever, are students asked to:
>   (a) formulate a question to be investigated;
>   (b) formulate an hypothesis to be tested;
>   (c) predict experimental results;
>   (d) work according to their own design;
>   (e) formulate new questions based on the investigation;
>   (f) apply an experimental technique based on the investigation just performed;
> (3) students are often asked to perform a variety of manipulative and observational procedures and to interpret the results of their investigations. Limited attention is given to planning and design questions and only limited opportunities are provided for higher level activities. The students commonly work as technicians following explicit instructions and concentrating on the development of lower level skills.

The 1978 U.S. use of the laboratory method has been highlighted and it has been shown that improvements are possible.

The main findings from research studies done on the laboratory method were the lack of consistent results. It seems clear that the laboratory method has a role to play in the arsenal of teaching methods, but it is also clear that many of the claims of the proponents of the laboratory method are not backed up by research findings.

2.4.3 The expository method

The expository method is perhaps most used in teaching science (Good and Brophy, 1978, 350). Perhaps it is also the one method science teachers are most familiar with (Good and Brophy, 1978, 356).

Aim and definition of the expository method

The main thrust of this method is to leave no doubt in the students' minds as to how to interpret the concept under consideration. The lecture or lecture-demonstration methods are examples. In its most ideal form the expository method allows no student questions because if taught properly all possible and potential problems are covered by the teacher. Hence, questions should not be neccessary, because there should be no uncertainties in the students' minds. In practice, both students and teachers exchange questions, but always with the aim of clarifying concepts or procedures or testing understanding.

In the section headed *2.4.1 Brief definitions of four science teaching methods* (see page 13) the expository method was defined as follows: by stating the concept, illustrating and rehearsing with examples, the students are told how to interpret concepts (Babikian, 1971, 201). The definition leaves no doubt that the onus is on the teacher to acquaint students with the pedagogical objective. The student is likened to a container into which the knowledge is to be poured.

In general, the concept is stated, illustrated and rehearsed with examples (Babikian, 1971, 201). The main media are voice (lecture) and the written word (chalkboard, whiteboard or overhead projector). If the method is combined with demonstration, or even laboratory, the students are told what effects to look for and how to interpret their observations. Even if audio-visual aids are used, the effects are always treated deductively, that is of a verification nature, and in a passive nature (Coulter, 1966, 185; Ramsey and Howe, 1969, 75; Zingaro and Collette in Bates 1978, 62).

Results of research studies

In this section the results of research studies on expository methods are presented. The expository method is compared with other methods like the laboratory and the discovery methods.

Studies which show advantages in the expository method

Babikian (1971, 208), in a study which compared the expository, laboratory and discovery methods, concluded that 'the investigation provided enough evidence to recognize the importance of expository methods in science education'.

Studies which show disadvantages in the expository method

Ramsey and Howe (1969, 78) and Good and Brophy (1987, 22-23) reported that most teacher-student interactions were shown to be taking place at a low cognitive level with 75% of the time being devoted to teacher talk. A high percentage teacher talk was characteristic of the expository method. The low cognitive level at which this discourse took place was not necessarily inherent in the expository method but it appeared that in practice this occurred.

Sorenson's findings (cited in Bates, 1978, 59) did not mention overall achievement, but indicated that the laboratory method achieved significantly better gains in critical thinking, understanding science and openmindedness when compared to the expository, lecture-demonstration, method.

Studies which show ambiguous conclusions

Oliver (cited in Bates, 1978, 58) compared three amended forms of the expository method: lecture-discussion, lecture-discussion with demonstration and lecture-discussion with demonstration and laboratory. He reported that although the lecture-discussion with demonstrations was significantly more effective during the first semester than the other two, by the end of the year no significant differences were found on recall of facts and theories. This conclusion agreed with the results of a study by Yager, Engen and Snider (1969, 85) who found no differences between laboratory, demonstration and discussion methods.

Shymansky (1972, xiv) and also Sherman (1989, 55), reporting on content achievement (recall of facts and theories), found no differences between lecture-demonstration and laboratory methods.

Concluding remarks on the expository method

From the collective impact of the results of research studies which compared the different methods of teaching science, such as the laboratory, expository and discovery methods, it appears that although the expository method has a role to play in science teaching, it is not more effective than other methods. In some areas, like content achievement, it is as effective as other methods. In areas like attitude development, it is not as effective. It appears that its high general use (Ramsey and Howe, 1969, 78; Good and Brophy, 1978, 350), at the expense or exclusion of other methods, bears no relation to its relative effectiveness.

2.4.4 *The discovery method*

Whereas in the expository method the students are left in no doubt as to how to interpret results, in the discovery method the student is the one who must put forward explanations and interpretations. Immediately, the contrast to the expository method is apparent: in the expository method the student is a receiver of knowledge, in the discovery method the student has to seek the knowledge.

Aim and definition of the discovery method

In its purest form the discovery method merely presents the students with materials, the students are expected to manipulate the materials until they are acquainted with the variables. They then start to control them. An hypothesis may occur to the student prior to a procedure or afterwards, even after results have been collected. Either way it is the conclusion that is all important in which the concept is isolated, understood and explained. Babikian (1971, 202) believes the teacher provides procedural instructions, that is the teacher may explain the procedure but not the concept. The teacher acts purely as a guide.

Rather than the students being totally responsible for discovering a conclusion through the control of variables Babikian's approach is a more pragmatic interpretation of the pure form of the discovery method. Babikian's interpretation Rotheram (1987, 632) would call guided discovery: '...(the student) follows detailed instructions to perform an experiment and is then expected to make conclusions'. It can be seen that although similar to the laboratory method, the discovery method concentrates on the results of the experiment, or investigation, not on the pathway towards the results. The term guided inquiry (Igelsrud and Leonard, 1988, 305) has also been used in which the student is supplied with an introduction, materials, a procedure giving directions but not a step by step guide, and a series of discussion questions.

Clearly, this more closely approaches the laboratory method.

Shulman and Tamir (1973, 1111) pointed out that the label discovery has different meanings. Learning to discover, discovery of content and learning by discovery refer to a process, an end goal and a mode of instruction respectively as they relate to the student, the subject and the teacher respectively. The discovery method used here refers to the mode of instruction.

The Learning Cycle proposed by Purser and Renner, a variation of a Learning Cycle proposed by Karplus and Thier (Abraham and Renner, 1986, 121; Igelsrud and Leonard, 1988, 304), is an adaptation of the discovery method in its pure form. The cycle contains three phases: exploration of the phenomenon, student invention of the concept and student application of the concept. Igelsrud and Leonard (1988, 305) believed that 'as support for discovery learning increases in research literature, strategies such as the Learning Cycle will be adapted to commercial biology lab programs'.

Results of research studies

Studies which show advantages in the discovery method

Ivins (1985, 115) showed that '...students experienced greater achievement and retention when directed discovery learning laboratories were used to introduce new concepts compared to the same laboratory activities used as verification laboratories'. Mulopo and Fowler (1987, 218) found that formal reasoners did significantly better in the discovery method and that attitude scores were higher in discovery groups compared with traditional groups.

Renner (1976, 222) believed that the reason physics was unpopular was that only formal reasoners had success. The probable reason for the success of formal reasoners is related to the necessity for students to draw conclusions from results. To obtain clear results the control of variables is essential. Wollman and Lawson (1977, 68) observed that '...young subjects do not have the capacity to keep track of variables if there are more than a few'. Lawson, Blake and Nordland (1975, 394) made an effort to teach students to control variables. They reported that '...the training was not successful in promoting...understanding of the controlling variables concept' in line with other studies that found specific concepts can be taught but transfer of those concepts is hardly ever obtained (Lawson, Blake and Nordland, 1975, 395).

Studies which show disadvantages in the discovery method

Mulopo and Fowler (1987, 218) found that although formal reasoners did significantly better in the discovery method, in general, achievement scores were lower in discovery groups compared to traditional groups.

Babikian (1971, 208) reported that the discovery method produced significantly poorer achievement results compared to laboratory and expository methods. Although Babikian blamed the poor results on students' lack of prior experience with the discovery method. The pure discovery method does not supply a procedure and students, generally, found it difficult to make a start (Pickering, 1985, 875).

Studies which show ambiguous conclusions

Whenever the discovery method was compared directly with another method (Babikian, 1971, 208; Gallagher, 1987, 353; Mulopo and Fowler, 1987, 218) conclusions appeared to be either in favour or against the discovery method. A conclusion of no significant differences was seldom found.

Concluding remarks on the discovery method

From the results of the research studies it would appear that the discovery method presents the most difficulties of the three methods discussed so far: the laboratory, expository and discovery methods. The difficulties are not limited to the students, but also apply to teachers. Therefore, while many of its proponents quote in theoretical terms what the discovery method aims to do in practice this method presents real problems in the classroom. It is not surprising, therefore, that this method is the least used of all method. For instance, Good and Brophy (1978) did not mention the discovery method once, neither in fact nor in spirit.

2.4.5 The inquiry method

Perhaps more than any other method the inquiry method has been open to many different interpretations. Although the value of the inquiry method is acknowledged in many curricula Welch et al. (1981, 37) reported that, although some aims were couched in inquiry terms, very little inquiry teaching took place. An overview of the different interpretations and approaches, which flow from these interpretations, is the subject of the following section.

Definitions and meanings of inquiry

According to Seymour et al. (1974, 349) one definition of inquiry is 'a set of activities directed towards solving an open number of related problems in which the student has as his principal focus a productive enterprise leading to increased understanding and application'. The inclusion of a set of activities, problem-solving, openness, enterprise, understanding and application, immediately tie in a number of related methods such as laboratory, problem-solving, discovery, the learning cycle. Peterson's definition (1978, 153) also includes related methods: 'scientific inquiry is defined as a systematic and investigative activity with the purpose of uncovering and describing relationships among objects and events'. Again, the discovery method is implicated.

The definitions of inquiry all hint at systematic experimentation; this is perhaps another source of student difficulty. Lawson, Blake and Nordland (1975, 395) found that most students lack the ability to control variables and, even more surprising, that this ability is difficult to teach. Most specific training attempts (by "specific" is meant: particular student experiments) were successful to some extent, but transfer of the concept of controlling all but one variable has never been successful. Added to this is the students apparent inability to think about manipulating data at the same time as thinking about theory (Pickering, 1985, 874). A consequence of student inability to control variables is that students exposed to traditional, that is verification laboratory exercises, immediately fall back on rigid procedures if they are confronted with non-explicit procedures (Pickering and Crabtree, 1978, 487).

There is a clear difference between discovery and inquiry. Discovery and inquiry have different focii, as Schwab (1966, 67) explained: 'students cannot be expected...to know what to look for...what questions to ask...this is the first and major responsibility of the teacher...his role (the teacher's) is to teach the student how to learn'. Discovery focuses on the end point of the investigation, inquiry focuses on the process used to get to the end point. Thus, inquiry, incidentally Schwab prefers the spelling enquiry, is characterized '...by the use of orderly and repeatable processes, reduction of the object of investigation to most simple scale and form, and use of logical frameworks for explanation and prediction' (Peterson, 1978, 153). These characteristics have caused a tendency to reduce inquiry to a method of doing controlled experimentation and to separate content from method (Schwab, 1966, 102). The latter is the very opposite of what the inquiry method is trying to achieve: content must always be seen in the context of how it was derived.

The inquiry method does not stress the accumulation of authoritative knowledge. It is more concerned with students discovering how scientists come to know what they know. Inquiry is defined as a search for knowledge or truth. The emphasis is on the search rather than the product (Sund and Trowbridge, 1967, 37). The inquiry method tries to overcome students' tendencies to give the first answer that comes to their mind. Wollman and Lawson (1977, 68) found most questions asked by teachers and parents were of low cognitive level and simplistic explanations and minimal responses were reinforced. The role model many teachers gave was found to be of someone having instant and correct answers. Perhaps it was the students' lack of experience (Babikian, 1971, 208) as well as the role model and reinforcement of recall and convergent questions and the lack of waiting time before accepting an answer (Wollman and Lawson, 1977, 68) which were the reasons why teachers and students found the inquiry method difficult (Igelsrud and Leonard, 1988, 305).

There are, broadly speaking, three different meanings attached to inquiry:

(1) the pedagogical method of teaching science by the inquiry method;
(2) teaching students the different methods that are used in science, that is the different methods of inquiry;
(3) science as conducted by scientists using the method of inquiry.

In this section these three interpretations are developed more fully.

The definitions of inquiry, given earlier in this section (see page 30), only hint at the different underlying meanings. Teaching and learning by inquiry versus science as inquiry are two meanings elicited by Tamir (1985, 88). The former refers to the type of interactions and activities that take place in the classroom and the method the teacher uses to promote learning. The latter refers to acquainting students with how science arrived at its conclusions or its knowledge. Science as inquiry can be taught by many methods, including the inquiry method. Schwab (Tamir, 1985, 89) argued that science as inquiry ought to be given more emphasis than the teaching method, but Tamir (1985, 93) believed that only through the inquiry teaching method do students

'...acquire a realistic image of science (as inquiry)'. Shulman and Tamir (1973, 1113) believed it was the phrase: teaching of science as inquiry, that confused the two meanings, but they stated that as long as the students are acquainted with science as inquiry the confusion was unimportant. They reported (1973, 1112) that more than 80% of programs labelled inquiry actually did very little teaching by inquiry method.

Rutherford (1968, 265) acknowledged Tamir's (1985, 88) two interpretations. Rutherford went further by concluding that by considering content and inquiry as one, because one has little meaning without the other, students could be taught the process of inquiry without it being absolutely necessary to use the inquiry pedagogical technique to gain understanding of inquiry as content. Rutherford argued that while the laboratory could be used to provide the student with experience in and knowledge of some aspects or components of the investigative techniques used in a particular branch of science the laboratory should only be used after careful analysis of the appropriateness of the experiment has been conducted. Four years earlier, Rutherford (1964, 84) argued along similar lines but proposed to devise laboratory experiences to simulate scientific inquiry as it actually happened, that is without the hindsight of theory known now.

A third meaning of the inquiry method strongly emphasizes knowledge and was suggested by J. R. Suchman as reported by Fish and Goldmark (1968, 263). Inquiry into science subject matter knowledge has students acting as scientists and inquire into theories and methods.

<u>*General forms of inquiry teaching*</u>

There is a limited consensus on what inquiry means and what scientists do. Summarizing this consensus Rotheram (1987, 632) proposed five options from which a teacher can choose when using the inquiry teaching method:

(1) 'guided discovery' (a student) follows detailed instructions to perform an experiment and is then expected to make conclusions;
(2) 'guided enquiry' involves the pupil more in experimental design and involves teacher-pupil discussions or may be structured through a series of written questions;
(3) 'guided exploration' involves the pupil in independent experimental design, using a flowchart as a generalised problem-solving strategy;
(4) 'free enquiry' also involves independent experimental design, but no form of assistance is provided;
(5) 'project' work should involve a series of enquiries and be mainly independent'.

These five options were developed from the three different meanings presented in the previous section. In terms of pedagogy the teacher gives the students increasingly more control from option one to option five. In the first three options the teacher can expose the students to different scientific methods. However, in the last two options students have to find the methods themselves. In terms of the third meaning of inquiry: inquiry into subject matter, in all five options the students must concentrate on knowledge even at the expense of variety or correctness of scientific method should time or circumstances dictate this.

The role of the laboratory in inquiry courses

Obviously, the lack of higher order cognitive skills, as found by Tamir and Lunetta (1981, 477), was quite a condemnation of the use of laboratory investigations. To convert a traditional laboratory into an inquiry laboratory Schwab (1966, 52), the most important of the inquiry proponents, advocated two changes: firstly, the main part of the laboratory activity should precede classroom discussion, secondly, the demonstration aspect should be inferior to both '...replacement of illustrations only of conclusions by illustrations of problem situations' (Schwab, 1966, 54) and to providing opportunities to do realistic but appropriately scaled-down scientific inquiry. Above all, there should be no artificial distinction between theory work and laboratory work.

Welch et al. (1981, 45) made two main recommendations in order to introduce inquiry teaching:

(1) personalize inquiry goals to what is appropriate to the student and school (not all inquiry goals need to be achieved);
(2) develop instructional resources and techniques; this development is the major task and the one that needs most attention.

Schwab (1966, 61) also regarded doubt as the key to a scientist's work and students should become familiar with a feeling of doubt. Note that doubt does not imply lack of precision or accuracy.

The question that remains is: why is the laboratory so important to inquiry teaching? Rubin and Tamir (1988, 477) believed that the laboratory is the '...core of the science learning process...(to) help the students understand important science concepts, provide concrete experiences...(to) develop process inquiry skills such as formulating problems and hypotheses, designing experiments, performing observations and measurements and drawing conclusions'. Unfortunately, this belief is not widely substantiated by research results as the next section shows.

Results of research studies

Studies which show advantages in the inquiry method

Rotheram (1987, 632-634), in an effort to make inquiry teaching easier for students to cope with, designed two flowcharts to help students with project work, the most demanding form of inquiry teaching. One flowchart was used initially to help students gain skills in inquiry. The second flowchart was designed to help students make a start on project work and was given to the student upon completion of the first flowchart. While Rotheram reported the flowcharts gave students confidence, he noted it was very time consuming and he did not perform a controlled experiment to get a statistical measure of the success of these flowcharts.

Guided inquiry was a term coined by Igelsrud and Leonard (1988, 305) as being a representative teaching strategy for the inquiry teaching method. Their method followed the definition of guided inquiry as outlined by Rotheram (1987, 632). No mention of the origin of the proposed strategy was made and no acknowledgement to Rotheram was given. The second of Rotheram's flowcharts was developed in 1984 (Rotheram, 1984, 660) in which he did claim to have statistical evidence of its success in a comparative test with guided discovery. In the 1984 study, no initial flowchart was used but more elaborate flowcharts followed the student progressively through a course of work; these latter flowcharts were not mentioned in the 1987 article.

Rubin and Tamir (1988, 477), to facilitate transfer to formal operation thought processes, applied Ausubel's idea of advanced organizers in a manner to make investigations interesting, applicable to the student's everyday life and avoid intellectual overload. A statistical analysis revealed that this aid was especially effective for weaker students but that the most difficult concepts for both control and treatment groups were: the need for a control and formulating hypotheses.

Renner and Lawson (Lawson, 1985, 602) found inquiry teaching produced significantly greater gains than traditional teaching methods at the college freshmen level. Furthermore, Renner and Paske (1977, 851) reported that inquiry teaching was more effective at producing reasoning gains than traditional teaching methods for concrete operational students. Lawson (1985, 604) also reported that concrete-based instruction appeared to be of benefit to all students.

Despite the barriers to inquiry teaching, many studies (Scott, 1966; Case and Fry, 1973; Seymour et al., 1974; Peterson, 1978; Mills, 1981; Pickering, 1985; Friedler and Tamir, 1986; Janners, 1988) have been conducted in an effort to either construct a workable program or determine the inquiry teaching's relative effectiveness. All studies reported success to some degree. Success was not limited to particular abilities and concrete experiences were found valuable. Peterson (1978, 158) noted that not all inquiry process skills, such as observing, questioning, formulating hypotheses, designing experiments, were equivalent in difficulty.

On the whole, all studies reported that, after initial hesitation, most students responded well, although several studies lacked rigorous statistical evidence to prove this.

Studies which show disadvantages in the inquiry method

Renner and Paske (1977, 851) reported that traditional methods were more effective than inquiry methods for transitional and formal reasoners. McKinnon (1970, 72) was surprised to find that out of 131 college freshmen only 22% could be classified formal operational, 51% were concrete operational and the rest transitional. At grade ten high school level Renner and cate (1985) determeined only 27% of students showed formal thought and Shemesh and Lazarowitz (1985, 16) found that in grade eleven 50% of students were at the concrete operational stage.

Students have been found to be incapable of controlling variables if there are more than just a few (Wollman and Lawson, 1977, 68) and incapable of constructing hypotheses and designing experiments when the experiences were somewhat unfamiliar (Welch et al., 1981, 39). These particular skills are very important in the inquiry method and this could explain Renner and Paske's (1977, 858) conclusions. Furthermore, it appeared that any inquiry process could be taught in a specific context but that transfer was a problem (Scott, 1966; Case and Fry, 1973; Seymour et al., 1974; Peterson, 1978; Mills, 1981; Pickering, 1985; Friedler and Tamir, 1986; Janners, 1988).

Both teachers and students experienced difficulties with the inquiry teaching method in a study performed by Igelsrud and Leonard (1988, 305). Case, Linn and Leone have all independently argued, in a report compiled by Rubin and Tamir (1988, 477), that '...inquiry-oriented laboratories often impose an overload on the short term memory of students who at the same time need to attend to new subject matter concepts, unfamiliar apparatus and problem solving tasks'.

Before inquiry teaching can be inducted into the curriculum there are more than the ten problems (Costenson and Lawson, 1986, 151), listed in *Concluding remarks on the laboratory method* (see page 22) pertaining to the 1978 U.S. state of the use of laboratories in high schools , to overcome. These ten problems could be overcome by teacher training and by providing role models for teachers. Schwab (1966, 63) believed the problems went much deeper, he listed the following problems:

(1) the way textbooks were published (although some new textbooks, for example Wilkinson and Nash's World of Chemistry and Nash and Hargraves' Chemistry Activities, now have minor components of inquiry slotted into them, with or without open-ended problems);
(2) students competed among themselves to gain better marks;
(3) the style of examination and testing used at all levels.

It is not surprising then that a study in the U.S.A. (Welch et al., 1981, 37) reported that, although some aims were couched in inquiry terms, very little inquiry teaching took place. The problems were felt to be lack of teacher training, teacher management problems, feared discipline problems, only for the brightest students and teachers felt that their primary purpose was to prepare students for the next level of schooling.

Another source of difficulty with the inquiry method of teaching science could be the lack of familiarity the students have with the method. This was one of the reasons suggested by Babikian (1971, 208) for the failure of the discovery method.

Lack of resources was a problem for Seymour et al. (1974, 351). The report by Seymour et al. (1974, 350) also mentioned that the inquiry method was considerably more time consuming compared to traditional methods.

Studies which show ambiguous conclusions

As was the case for the discovery method (refer page 29), it appears that conclusions reached by research studies were either in favour of the inquiry method or against it.

Concluding remarks

Researchers agree on what the skills involved in the inquiry process are: observing and measuring, problem identification, interpreting data, generalizing, formulating hypotheses, designing and evaluating experiments (Welch et al. 1981, 34; Peterson, 1978, 158). However, it appears that they rarely agree on how to teach it. Indeed, research results are often difficult to compare because the methods of inquiry teaching have been so different. Nevertheless, all studies have some positive results to report despite a lack of resources and a greater demand on both teachers and students.

Specific teaching examples used in this study are given in section 4.7 in acknowledgement of the problem of comparing different inquiry methods. The units taught are presented in full in Appendix F. To allow for lack of student ability in controlling variables (Lawson, Blake and Nordland, 1975, 395) the inquiry method used in this study starts closed-ended and gradually presents more open-ended tasks. This approach progressively gives students experience (Babikian, 1971, 208) and overcomes students' perceptions of the level of difficulty of the inquiry method. This perceived level of difficulty was encounterd by Igelsrud and Leonard (1988, 305). The success of this approach was demonstrated by comparing inquiry students' achievement scores with students taught by traditional methods (refer table 6.5). The inquiry method was not superior in achievement scores on all criteria but produced at least equal results, in line with the results of many other studies (Peterson, 1978; Mills, 1981; Pickering, 1985; Friedler and Tamir, 1986; Janners, 1988).

Peterson (1978, 158) noted that the inquiry process skills were unequal in difficulty. The results of this study agree with Peterson: all students found experimental methodology particularly difficult (refer: implications for science teachers, page 184). This study also confirmed Seymour's et al. (1974, 351) observations that the inquiry method is more time consuming compared with traditional science teaching methods. However, on the whole, this study found that syllabus content does not need to be changed in order to teach by the inquiry method.

The literature review has clarified what the laboratory, expository, discovery and inquiry methods entail. This information has been used in chapter four, the design of the treatment method (the inquiry method of teaching science), to ensure similarities in treatment used in other studies done on the inquiry method. The literature review also showed what data on student characteristics could be gathered, such as cognitive style, cognitive preference, attitude and student reasoning ability. The instruments used to collect these data are described in the next chapter: instruments used to describe students.

CHAPTER 3

INSTRUMENTS USED TO DESCRIBE STUDENTS

In this thesis four different types of instruments were used to describe the students who participated in the project. These four instruments were selected because they had previously been used in inquiry related research. They offered useful potential measurements of student characteristics on which students might be separated with regard to relative success in the inquiry method of teaching. A fifth instrument was developed during the course of this research. This instrument has not been used in previous research.

This chapter is concerned with the description of the five instruments and brings forward other relevant research literature. Each instrument is dealt with in turn:

(1) Cognitive Style of Categorization Behaviour;
(2) Combined Cognitive Preference Inventory;

(3) Test Of Science Related Attitudes;
(4) Structures of Observed Learning Outcomes;
(5) the Student Perceived Characteristics for Success.

3.1 Cognitive Style of Categorisation Behaviour

This instrument is related to a student's cognitive style of categorisation behaviour. The principles underlying the construction of the instrument are explained first and previous use of the instrument is discussed. The details of the construction are given because the instrument was developed as part of this thesis. Results pertaining to pilot testing are also described.

The instrument used in this thesis to measure cognitive style is based on an instrument used extensively by Scott (1964, 1970, 1973). This was devised by Sigel, the Sigel Cognitive Styles Task (Scott, 1970, 95), and described students in the way they responded to a course in science instruction using the inquiry method of teaching.

3.1.1 Cognitive style

In order to understand what is meant by cognitive style it is useful to consider how humans organize objects, or things in general, through the use of language. The following examples illustrate this organization. Upon entering a room one may give items particular names, for example chair, table or mat. When considering the house as a whole, chairs and tables may be classified as furniture and mats as floor-coverings. The house itself may be described as Georgian, dilapidated or beautiful. Each of these descriptions focus on different aspects: style, structure and appearance, respectively. Each individual approaches the categorization of objects differently by focussing on different aspects of the objects. Sigel found that 'when the responses of a variety of individuals are observed, specific patterns of preference emerge...an individual's "style" of categorisation can be generally associated with one of several of six broad categories of verbal labels' (Scott, 1970, 97). Thus a person's cognitive style of categorisation behaviour is dependent on '...the kinds of cues a person uses in perceiving similarities and, subsequently, in categorizing the various dimensions of his environment' (Scott, 1973, 323). Scott used Sigel's definition of cognitive style: '...Style is an umbrella term that includes a variety of processes...It refers to mode(s) an individual employs in perceiving, organizing and labelling various dimensions of the environment' (Scott, 1964, 7).

3.1.2 Details of the six cognitive styles

The six broad categories of Scott (1964, 9), alluded to previously, have awkward sounding titles but are easily understood using examples:

(1) "descriptive, Part-Whole" (P) is the label given to a categorization where only part of the whole observable object is used, such as the wheels of a car;
(2) "descriptive, Whole" (W) is attached to a whole item, such as in the description of a car the whole car is red or that it is a car;
(3) "Relational-Contextual" refers to a label describing the object's interdependence with another object, such as the car can tow a trailer. This label is also known as "Functional-Interdependence" (FI);
(4) "Categorical-Functional" (CF) is the category reserved for labels that denote the use of the object, for instance the car can be used to drive to work;
(5) "Categorical-Class naming" (CC) here the object is seen as belonging to a wider group of similar objects, for example the car is seen as a passenger vehicle or as a method of transportation;
(6) "Categorical-Inferential" (CI) describes some inferred attribute assigned to the object that may not be actually seen in the object but may come from experience or imagination, such as the car is beautiful or looks fast or sporty (note that describing it as a sports-car would be

designated Categorical-Class naming).

There are, of course, many student responses to the instrument items that do not fit neatly into a specific category. For instance, when a student responds to the stimulus of a car and describes it as metallic, as opposed to plastic, is it being viewed as a whole or did the student focus on just a part of the whole? In order to ensure consistency during the marking of the instruments, the author kept notes on which category a particular answer was assigned to; these notes are presented in Appendix A.

3.1.3 Results of research studies on cognitive styles

Before explaining the way instrument items were constructed some general comments pertaining to cognitive style of categorization are appropriate.

Kagan, Moss and Sigel (cited in Scott, 1973, 323) and Ogunyemi (1973, 59) found that as students develop mentally there is a shift away from wholistic descriptions to more analytical categorisation. Presumably, what they mean by this is a more concentrated focus on details that place the object in certain classes, that is a focus on similar and dissimilar attributes and properties of the objects. This shift was found to be a slow, steady, continuous one and '...a student generally retains his stylistic position relative to other students...Thus, a subject's cognitive style seems to be somewhat resistant to changes...and efforts to train students to become more analytical had not met with success' (Scott, 1973, 323).

Longitudinal studies at high school level (Scott, 1973, 326) have revealed that with age students use more Part-Whole descriptors but do not change significantly in the other five categories. Scott (1970, 100) also noted a decrease in use of Categorical-Inferential labels, although the level of significance is not clear. Using students from grades six and seven Scott (1964, 14) found that high inductive reasoners (able to draw conclusions from results) tend to score high in the Categorical-Inferential category and that for males the use of the Categorical-Functional label decreases with age. In studies researching the effects of inquiry training, it was found that inquiry students were more fluent and flexible in their labelling (Scott, 1970, 95) and paid more attention to Part-Whole descriptors (Scott, 1970, 99; Scott, 1973, 326) than did non-inquiry students.

3.1.4 Instrument item construction

For the current study the stimuli were presented in similar ways to the method developed by Dr. Irving Sigel (Scott, 1964, 9). The original method used cards to present pictures to individual students. In this study names of the pictures were used in place of actual pictures. This method relied on the students' familiarity with the names to form a picture in their own minds. This was done to save both time and paper because the draft instrument contained 30 items.

For each item in the instrument the student was presented with three objects. The objects were all related in some way. The students were asked to select any two of the three objects and write down how the two selected objects were related. Prior to starting the instrument brief instructions with three example answers were read. In the examples, answers belonging to each of the six categories were supplied and it was stressed that no answer was right or wrong. Even if the answer was obviously untrue as long as the student believed it was true, the answer was acceptable. All 30 items had three objects associated with them. They were selected so as to be totally familiar to all students and included a spoon, fork, car, truck, nurse, carpet, cat and icecream. An example of an item read out to students prior to the test is:

1. book, magazine, newspaper- reasons for selection (=*):

   * * both have separate articles;
   both are serials;
   an article in a newspaper may be

expanded on in the magazine;
both are thrown out after reading;
they are both thin;
both are meant to entertain;
they are both paperbacks.

3.1.5 The pilot instrument

The pilot instrument was administered to three different classes of vastly different maturity and background in order to obtain variance of answers, to ensure instrument appropriateness and check completion time. A grade three from a city primary school, a grade nine from a country school and a grade eight from a city school were used. A total of 78 students completed the pilot instrument. The students had an unlimited time to answer the items. Students simply selected two out of the three objects and indicated this by circling names. To the right of the names of the three objects a space was provided for the students to write a brief statement describing how the two items were related. Each statement was then classified into the appropriate category.

When all statements had been classified into categories the number of responses in each category were tallied. A profile was obtained, for each student, showing the category with the highest tally, the next highest tally, and so on.

The responses to the pilot instrument were used to reject items that were inconsistent. Instrument items that extracted inconsistent answers from students belonging to the same type of cognitive style were rejected. This was done in the following way:

(1) one point was assigned for each classifiable response;
(2) students with a difference of four or more points between the highest and the next highest categories of categorisation were selected;

(3) the students from (2) were classified by their highest category score;

(4) these selected students were grouped by category;

(5) those items which less than 75% of the group did not select as belonging to their highest category were rejected as inconsistent.

The reason for this procedure was to ensure that if students was strongly oriented toward a particular style of categorisation there was more than 75% chance that they would respond to that item in a way consistent with his or her dominant style.

In total, eleven items were rejected, leaving a total of nineteen items. This was considered a sufficient number for reliability calculations.

3.1.6 Student classification into cognitive styles

To give fair weighting to each style raw scores were converted to percentages. This was done because not all students completed all 30 items. Those with less than 60% completion of items were not used in the calculations. A t-test (Pilliner, 1975, 17) of significant difference between highest and next highest score difference of a minimum of four was conducted. This test was performed on all students in grade eight. The grade eight group t-test result was highly significant ($p< 0.001$) (Roscoe, 1975, 429). Analysing only those scores for which the raw scores differed by a minimum of two, the difference between highest and next highest score was still significant ($p< 0.004$).

The consequence of this was that it was possible to classify a student with great confidence in one particular style of categorisation if the raw scores for different styles differed by more than two. If the raw scores differed by two, it was statistically still justifiable to classify the student by the highest score. If the raw scores differed by only one, then a singular classification was unacceptable and a dual, or triple, classification was required. This is summarized in table 3.1.

Table 3.1 Determining when a student can be classified by a singular, dual or triple category based on the difference between the student's highest and next highest scores.

Difference (d) between highest and next highest scores.	Consequence with regard to classification of student
d=0	dual or triple classification
d=1	dual or triple classification
d=2	classify by highest score
d>2	classify by highest score

3.1.7 Analysis of the pilot instrument

The data from the grade nine sample was tested in the same way. The grade nine t-test result was highly significant ($p < 0.0005$) for a difference of two between the highest and the next highest classification styles' scores. This result indicates that data trends between year groups can safely be analysed for changing shifts in patterns of styles in categorisation. The analysis was performed in three steps:

(1) comparison between grades three and nine;
(2) comparison between grades nine and eight;
(3) comparison between grades eight and three.

The results of this analysis are displayed in tables 3.2, 3.3 and 3.4.

Table 3.2 A comparison of the cognitive styles of grades 9 and 3.

Cognitive style	Grade emphasis on category	Relative difference
Whole	9 > 3	most difference
Part-Whole	9 < 3	
Categorical-Inferential	9 < 3	
Categorical-Class	9 < 3	
Categorical-Functional	9 > 3	
Functional-Interdependence	9 = 3	least difference

Note: multiple t-test (Roscoe, 1975, 429) showed significant difference ($p<0.005$) between grades 9 and 3.

Table 3.3 A comparison of the cognitive styles of grades 9 and 8.

Cognitive style	Grade emphasis on category
Whole	9 > 8
Part-Whole	9 > 8
Categorical-Inferential	9 < 8
Categorical-Class	9 < 8
Categorical-Functional	9 < 8
Functional-Interdependence	9 = 8

Note: multiple t-test (Roscoe, 1975, 429) showed significant difference ($p<0.008$) between grades 9 and 8.

Table 3.4 A comparison of the cognitive styles of grades 8 and 3

Cognitive style	Grade emphasis on category
Whole	8 < 3
Part-Whole	8 < 3
Categorical-Inferential	8 = 3
Categorical-Class	8 > 3
Categorical-Functional	8 > 3
Functional-Interdependence	8 = 3

Note: multiple t-test (Roscoe, 1975, 429) showed significant difference ($p<0.004$) between grades 8 and 3 when Categorical-Inferential and Functional-Interdependence were excluded..

In this section a discussion of the results is presented. First the grades three and nine were compared, in a multiple t-test, for a significant difference in styles. A significant difference of $p < 0.005$ (two-tailed) (Roscoe, 1975, 429) was found mainly in the styles: Whole, Part-Whole, Categorical-Inferential, Categorical-Class and Categorical-Functional in order of decreasing difference. Grade nine was more Wholistic, less concerned with Part of the Whole and also less concerned with Categorical-Inferential and Categorical-Class. They concentrated more on the use of items (Categorical-Functional) and viewed stimuli more on a Wholistic basis. It must be remembered that the grade nine sample was drawn from a rural population as opposed to the grade three that was drawn from a city primary school. In a tentative pre-emption on the discussion of the Combined Cognitive Preference Inventory, it was noted that the tendencies of the rural grade nine sample were in line with the observations made by Tamir and Kempa (1978, 150). They concluded that agricultural schools are predominantly application-oriented compared to occupational schools and city schools. City schools differed from both agricultural and occupational schools. Agricultural and occupational schools were largely similar. Presumably, it is logical to view stimuli as a Whole if its use is the main interest to the student.

Second, the grade nine was compared to the grade eight and tested for significant differences. Again, a significant difference was found: $p < 0.008$ (two-tailed) (Roscoe, 1975, 429). The grade nine sample rated higher in both Wholistic and Part-Whole styles; the grade eight was predominantly oriented in the Categorical-Functional style. The styles Categorical-Class and Categorical-Inferential also occurred more often in the grade eight.

Third, the main shifts from grade three to grade eight, both city schools, are from Wholistic and Part-Whole and toward Categorical-Functional and Categorical-Class ($p < 0.004$) when the styles Relational-Contextual and Categorical Inference are not considered).

These results agree with those published by Kagan, Moss and Sigel (cited in

Scott, 1973, 323 and in Ogunyemi, 1973, 59) in that there is a shift from the Wholistic style as students mature and the style becomes more analytical. In this research the shift was toward Categorical-Functional and Categorical-Class. The results do not run parallel with the findings of Scott (1973, 326); he found a trend toward Part-Whole and no significant change in the other styles as students matured.

The literature makes no mention of any studies done in which a comparison was made between abilities within the same year groups. It was decided to compare the top 25% with the bottom 25% of students based on academic performance of both the grade three and the grade eight samples. In both samples the differences were significant with $p< 0.02$ and $p< 0.001$ for grades three and eight, respectively. Even more interesting was the observation that the differences in cognitive styles were the same in both samples. In both cases the lower ability 25% placed less emphasis on descriptive Part-Whole, more emphasis on descriptive Whole and somewhat less emphasis on Categorical-Class and -Inference. Obviously, the lower ability students are less analytical and more Wholistic oriented in style. This conclusion runs parallel with the conclusion by Kagan, Moss and Sigel who based their conclusion on difference in maturity and not ability as was done in this study.

3.1.8 Reliability considerations

A nineteen item instrument was considered sufficient from a reliability point of view. This is despite the fact that shortening an instrument reduces reliability. If the instrument is too long students may become bored or fatigued and this reduces reliability (Cohen, 1976, 390).

Reliability is 'the extent to which a test gives consistent results' (Pilliner, 1975, 52). For this instrument the reliability was calculated using the Cronbach-Alpha method in line with calculations done by (Scott, 1964, 1970, 1973). The Cronbach-Alpha reliability of this instrument varied for different grades from

0.59 to 0.74, that is from moderately unreliable to moderately reliable. These values are similar to the split-half reliability calculations performed on the original instrument based on the twelve picture cards (Scott, 1973, 324). The values reported by Peterson (1978, 156) were 0.62 to 0.69. Peterson's (1978, 156) results were based on 15- and 11-item instruments, respectively, and the values were similar to those reported by Cooley and Klopfer, Suchman and Tamir (Peterson, 1978, 156).

3.1.9 Improvements that can be made to the instrument

This instrument is deemed satisfactory for the purpose of this thesis, but a number of steps could be taken to improve it for general use. The first priority should be the development of more items that promote the cognitive styles of Categorical-Inferential and, especially, Relational-Contextual (Functional-Interdependence). Very few students answered items in the latter style. An investigation could lead to the deletion of items that produced mostly one style of response by retaining only those items that give a wide range of response styles. Rather than allowing the students to write down a response that is unique to them, considerable time-saving could be achieved by providing six options and asking the students to select the answer they like best. This format would overcome the problem of classifying responses into category styles when the response appears to fit into more than one category. This would be an advantage with regard to reliability if there is more than one marker.

Finally, it would be interesting to find out how student responses change from one style to another if not just one response per item is required but the student has to think of three, or more, answers for the same item.

3.2 Combined Cognitive Preference Inventory

In this situation the term cognitive preference is used to describe the way an individual attends to scientific information. It is quite different from achievement in that achievement tests measure what students can do with the information or remember about it but gives no indication of how students viewed the information they were given. Cognitive preference instruments typically measure how the student intellectually processes the information. In 1964 Heath first used cognitive instruments. Heath explained cognitive instruments in the following way: 'the interest is not in whether the student can identify correct or incorrect information but rather in what he is likely to do with the information intellectually' (Heath, 1964,241).

3.2.1 Details of cognitive preference styles

Heath suggested that on the whole individuals attend to scientific information in four different modes (Heath, ibid; Tamir and Kempa, 1978, 143):

(1) acceptance of scientific information for its own sake, that is without consideration of its implication, application, or limitations. This mode is designated as Recall;
(2) acceptance of scientific information because it exemplifies or explains some fundamental scientific principle or relationship. This mode is designated as Principles;
(3) critical questioning of scientific information as regards its completeness, general validity, or limitations. This mode is designated as Questioning;
(4) acceptance of scientific information in view of its usefulness and applicability in a general, social, or scientific context. This mode is referred to as Application.

In the early versions of cognitive instruments students were required to select one style only in a multiple choice instrument that the student found most satisfying or appealing. In these versions the three options that were not selected were treated as equally unattractive. Later versions abandoned this approach and asked the student to order all four responses from most satisfying to least satisfying. Most satisfying earned four points, least satisfying earned one point. It was felt by Tamir and Kempa (1978, 144) that this graded approach contributed much more to the student's overall cognitive style. The graded approach developed a student's profile and showed which style is more preferred to another. Thus, in a twenty item instrument a maximum score of 80 is possible and a minimum of twenty, with an average score of 50 per cognitive mode.

3.2.2 Reliability of the cognitive preference instrument

In Tamir and Kempa's research (1978, 145), three instruments each of twenty items were used in Physics, Chemistry and Biology. The instruments were administered to several schools designated as city, agriculture and occupational. The Cronbach alpha reliability figures are presented in table 3.5.

Table 3.5 Cronbach alpha reliability data of several research studies.

Research study	Cronbach alpha reliability	Number of items
Tamir and Kempa (1978, 145)	0.58 - 0.83	3 separate tests 20 items each
Tamir (1975, 240)	0.73 - 0.84	20
Tamir (1976, 57)	0.81 - 0.90	40

3.2.3 Results of research studies on cognitive preference styles

Since Heath (1964), more than twenty-eight studies have been conducted on cognitive preference style (Tamir, 1978, 60; Tamir and Cohen, 1980; Carter, 1982; Mcnaught, 1982; Shuaibu and Ogunsola, 1983; Tamir, 1983; Tamir and Jungswirth, 1984; Okebukola and Jegede, 1988; Tamir, 1988). A further seven articles have been published which use previous studies to either argue alternative interpretations or to discuss validity and reliability concerns (Tamir, 1978; Van Den Berg, Lunetta and Tamir, 1978; Jungwirth, 1980; Van Den Berg, Lunetta and Tamir, 1982; Rost, 1983; Tamir, 1985; Gardner and Tamir, 1989). Many other studies have also been conducted researching into related but different cognitive style constructs such as: sensing, intuitive, judging and perceiving styles (Novak and Voss, 1981); factual and principal orientations (Mackay, 1972); competitive and cooperative styles (Behr and Eastman, 1978; Okebukola, 1986); field dependent and independent styles (Shymansky and Yore, 1980; Harpole, 1987).

Ipsative and normative are the two methods used in the studies related to measurement of student preferences. Ipsative refers to a multiple choice response for least preferred where the selection of one preference excludes the others. Normative methods require students to rank responses that correspond to four cognitive styles and provisions for tied ranks are made. The conclusions of the studies on cognitive style can be sorted into seven groups that relate to:

(1) grade;
(2) subject;
(3) school environment;
(4) achievement and cognitive ability;
(5) student characteristics;
(6) the Questioning-Recall and Principles-Application dichotomy;
(7) inquiry courses.

The research conclusions in each of the seven groups will be presented in turn.

(1) Grade

The higher the high school grade the more Principles was preferred as a cognitive style. The preferences for the cognitive styles Recall, Application and Questioning diminished in this order (Tamir and Kempa, 1978, 146).

(2) Subject

Tamir and Lunetta (1978, 63) concluded that preferences were related to both the science disciplines and to the topics within the disciplines. In Chemistry, Tamir (1975, 241) concluded that students polarized along two axes. One axis represented the cognitive styles combination low Recall/high Questioning versus high Recall/low Questioning. The other axis represented the combination low Application/high Principles versus high Application/low Principles. For the discipline Biology this polarisation was not apparent.

(3) School environment

Tamir (1975, 246) concluded that city schools were significantly different in cognitive styles from occupational and agricultural schools. City schools had higher preferences for Questioning and Principles than non-city schools. The latter had higher preferences for Application than city schools.

(4) Achievement and cognitive ability

Cognitive ability and cognitive preference were proven to be separate entities by Tamir (1976, 57; 1978, 62). Tamir and Kempa (1978, 150) concluded that the higher a student's achievement score, on end of unit tests, the higher that student's score on the cognitive style Principles was likely to be. In agreement with this conclusion Tamir (1978, 60) also found a significant relationship between cognitive preference profile and academic achievement. Two years earlier, Tamir (1976, 57) had concluded that high achievers had a high preference for Questioning, a weak preference for Principles and a strong dissatisfaction with Recall. Tamir and Lunetta (1978, 64) suggested that a high preference for Questioning and a low preference for Recall represented 'a very high level of intellectual curiosity and a desire to learn more'. Tamir

(1975, 241) suggested a high positive Questioning-Recall score indicated a keen student whereas a high negative Questioning-Recall score indicated a student lacking in the ambition to learn more.

(5) Student characteristics

Although some research had been conducted in this area (Tamir and Lunetta, 1978, 63) it was known that the sex of a student was not a significant variable in determining cognitive preference style (Tamir and Kempa, 1978, 150). However, Tamir (1975, 250) concluded that females are more affected than males by the nature of the curriculum, for instance an inquiry or traditional curriculum. Teacher bias toward the nature of the curriculum was also an important factor in determining the cognitive profiles of students (Tamir, 1975, 250). Tamir (1975, 258) also concluded that cognitive preferences were relatively stable over periods ranging from one semester to three years.

(6) The Questioning-Recall and Principles-Application dichotomy

Tamir and Kempa (1978, 150) concluded that two scales: curiosity and utility, would be more meaningful than the four cognitive styles separately. The curiosity scale was a derived scale. It was obtained by subtracting the Recall score from the Questioning score. The utility scale was a derived Principles-Application score. In support of this conclusion, Tamir (1975, 241; 1978, 62), using factor analysis, determined that a high preference for Questioning tended to predict a low preference for Recall and vice versa. This tendency was also reported by Tamir and Lunetta (1978, 64).

(7) Inquiry courses

Tamir (1975, 238) investigated the cognitive profiles of students participating in an inquiry oriented science course. The length of student exposure to the inquiry oriented course was not made clear, but Tamir did state that the students had been studying this course for several years. Results showed that the course caused high Questioning/low Recall preference profiles but not consistently. Apparently, it is more the Principles mode that seems to distinguish between inquiry and traditional oriented courses (Tamir, 1975, 238). Tamir and Lunetta (1978, 61) concluded that an inquiry oriented Biology course promoted a high Questioning/low Recall mode in high ability students.

3.3 Test Of Science Related Attitudes

3.3.1 The rationale for using this instrument

The third instrument used in this study to describe student characteristics was the Test Of Science Related Attitudes. It was an instrument devised by Barry J. Fraser (Khalili, 1987). The instrument was based on Klopfer's conceptual classification of affective aims for science education (Khalili, 1987, 128). This particular instrument was included primarily for two reasons. The first reason was due to the presence of one scale out of the seven scales in the instrument which measured attitude to inquiry (Fraser, 1978, 509). In particular this scale measured attitude to scientific experiments and scientific inquiry as ways of obtaining information about the natural world. Khalili (1987, 127) described attitude to science as the disposition of mind for or against science and scientists.

The second reason was the inclusion of the adoption of scientific attitude scale, which, according to Khalili (1987, 127) meant having a regard for evaluation, thoroughness and attention to detail. The distinction between scientific attitude and attitude to science was important as both these concepts were also used in the laboratory objectives as categorized by Shulman and Tamir (1973, 1119 and cited in Bates, 1978, 57).

3.3.2 Details of T.O.S.R.A.

Table 3.6 shows the seven scales of the instrument and sample items as they appear in the instrument.

Table 3.6 Scale name and Sample item for each T.O.S.R.A. scale.

Scale name	Sample item of the scale
Social implications of science	Scientific discoveries are doing more harm than good.
Normality of scientists	Scientists usually like to go to their laboratories when they have a day off.
Attitude toward scientific inquiry	I would prefer to find out why something happens by doing an experiment than by being told.
Adoption of scientific attitudes	I like to listen to people whose opinions are different from mine.
Enjoyment of science lessons	Science lessons bore me.
Leisure interest in science	I dislike reading newspaper articles about science.
Career interest in science	I would like to be a scientist when I leave school.

(Fraser, 1978, 510)

Students are asked to respond to 70 items on a five point Likert scale. There are, therefore, ten items for each scale, resulting in a minimum of ten and a maximum of fifty points per scale.

Fraser (1978, 514) found that intercorrelations between the seven scales varied from 0.10 to 0.59 with the highest occurring between three scales: Enjoyment of science lessons, Leisure interest and Career interest in science. Fraser concluded that all seven scales were acceptable despite the high figures.

Fraser (1978, 514) advocated the use of the instrument to determine the profiles of both groups and individuals. It is interesting to note that the three scales with the highest intercorrelation coefficients also had the highest standard deviations for all grades (Fraser, 1978, 513). No other patterns, neither on grade nor on scale bases, were discernible.

3.3.3 Reliability of the instrument

The instrument has been field tested for grades seven to ten in a variety of geographical and socio-economic environments in Australia (Fraser, 1978, 510-511). The Cronbach alpha reliability coefficient for each of the scales varied from 0.64 to 0.93 and test-retest reliability was good with an average of 0.78 (Fraser, 1978, 512).

Khalili (1987) tested Fraser's T.O.S.R.A. instrument in the U.S.A. for reliability and discriminant validity determination. Khalili (1987, 128-129) reported high validity and reliability coefficients of 0.69 to 0.91. The seven scales were, however, not distinct. The intercorrelation data between the three scales Enjoyment, Leisure and Career provided different values for different studies: Fraser (1978, 514) found 0.57, Schibeci and McGaw report a value of 0.73 and Khalili a value of 0.84 (Khalili, 1987, 130). The high correlation between the Enjoyment, Leisure and Career scales indicate the three scales measure the same thing. As a result the three scales can be collapsed into one scale. On the basis that the scale Enjoyment of science had the highest reliability of the three scales, this scale was used in this study to represent all three scales. Khalili (1987, 133-134) provided more justification for collapsing the three scales into one Enjoyment of science scale by showing that in a rotated varimax analysis all three scales fell on the same factor with the Enjoyment scale the most restricted to that factor. The scales Social implications and the Normality of scientists are not relevant to this study and were not used as bases for classifying students.

The three most applicable scales from the T.O.S.R.A. instrument were the Attitude, the Adoption and the Enjoyment scales. These formed part of the data gathering for this research.

3.4 Structures of Observed Learning Outcomes

3.4.1 The aim of the S.O.L.O. instrument

This instrument was developed and designed by Biggs and Collis in 1982 (Collis and Davey, 1984, 4). The main aim of the instrument is to classify and 'evaluate student responses to particular tasks' (Pallett, 1983, 4). In this regard it is not dissimilar to Piagetian tasks in the sense that the response of the student is classified and not the student. Once a profile of responses has been obtained the student is then classified as belonging to a particular stage of development. This classification gives the teacher the opportunity to determine the student's stage of cognitive development. Arguments have been presented to use S.O.L.O. as a theoretical framework on which to base assessments as part of the current school based curriculum developments (Biggs and Collis, 1989, 151) and criterion based assessment moves (Collis and Biggs, 1989, 25).

3.4.2 Details of the S.O.L.O. levels

Pallett (1983, 5) reported five levels into which the student responses can be classified: Pre-structural, Uni-structural, Multi-structural, Relational and Extended Abstract. These levels increase in level of abstraction, in the number of dimensions the student can organize the data and the degree of openness of the conclusions the student can reach from the data. Table 3.7 shows the relationship between the S.O.L.O. levels, Piagetian stage of cognitive development and response structure in terms of how the response is related to the data provided in the item.

Table 3.7 Cognitive level, level of involvement and response description relating Piagetian stage, S.O.L.O. level and student response (adapted from Fisher, 1986).

Piagetian stage	S.O.L.O. level	Capacity	Relating operation
Formal operation (16+ years)	Extended Abstract	maximal: cue+relevant data+inter-relations +hypothesis	Deduction and induction. Can generalize to situations not experienced
Concrete general-ization (13-15 years)	Relational	high: cue+relevant data+inter-relations	Induction. Can generalize within experienced context
Middle concrete (10-12 years)	Multi-structural	medium: cue+isolated relevant data	generalizes in few independent data
Early concrete (7-9 years)	Uni-structural	low: cue+ one relevant datum	concludes in terms of one datum
Pre-operation-al (4-6 years)	Pre-structural	minimal: cue+an ir-relevant datum	denial, tautology relevant response

Legend: +=as well as.

Each item in the S.O.L.O. instrument was constructed the same way. A stem provided information with several pieces of data (Pallett,1983,7), following the information five questions were asked one at each S.O.L.O. level. The question sometimes supplied more data, especially at the top two levels. The marker was provided with answers to each question and simply marked the answer correct or incorrect at each S.O.L.O. level, building up individual student profiles.

Tamir and Collis (1990), at the time of writing, were investigating possible relationships between cognitive preferences, S.O.L.O. level, school academic achievement and gender. They requested not to be quoted because the preliminary results were based on small student samples. They did note, however, that S.O.L.O. levels were related to ability and that cognitive preferences were related to the style and orientation of the student's information processing skills (Tamir and Collis, 1990, 3). A relationship was expected because Tamir, in 1985 (Tamir and Collis, 1990, 3), found that high achievers had a preference for Principles and Questioning and a low preference for Recall. It could be hypothesized that high achievers were in the top two levels of S.O.L.O. Hence, there should have been a relationship between preferences and S.O.L.O.

Collis and Davey (1984, 6) suggested the labelling of SOLO levels as I, S, M, R, G, to reflect the use of the data supplied in the stem in the student's response: Incidental, Single, Multiple, Related, General, respectively, to indicate Prestructural, Unistructural, Multistructural, Relational and Extended Abstract, respectively. In this thesis, therefore, the students' responses are classified as either Incidental, Singular, Multiple, Relational, General Abstract.

3.4.3 Reliability considerations

The validity and reliability of the S.O.L.O. instrument have been determined by Collis and Davey (1984). They determined that the construct validity was satisfactory, the coefficient of reproducibility was 0.85 and the Cureton's KR-20 reliability coefficient was 0.85 as well. These results were based on years seven and nine students (Collis and Davey, 1986, 660).

A Guttman scalogram analysis was performed on all S.O.L.O. items. The reliability of the instrument items used in this thesis varied from 0.80 to 0.97

at significance levels of 0.05 or less. These figures meant that the items used in this thesis were all statistically acceptable (Collis and Davey, 1986, 660). The scalogram investigated the possibility of reversals in the response patterns, meaning a student may answer a low level incorrectly yet a higher level correctly. Collis and Davey (1984, 14) expected these reversals to occur 'very infrequently (e.g. 2% of the time)'. Although they acknowledged that the level of difficulty of the items was not perfectly consistent for each of the Incidental, Singular, Multiple, Relational and General Abstract response levels (Collis and Davey, 1986, 660).

3.4.4 Results of research studies

Collis and Davey's study (1984, 15) revealed a gradual shift from lower to higher S.O.L.O. levels from grade seven to grade nine, with very few General Abstract rated students. The majority of grade nine students were at level Multiple (31.5%), with a slightly lower percentage at level Relational (23%). A later report gave the same results for grade nine; for grade seven the percentages were Multiple 22% and Relational 16% (Collis and Davey, 1986, 662). These latter percentages were, again, quoted by Collis and Biggs (1989, 18) and were most likely taken from the same study. The percentages have obvious teaching implications especially in subjects like Physics where a large number of concepts are abstract. Table 3.8 shows the percentage distribution of students at each S.O.L.O. level for the questions used in this study.

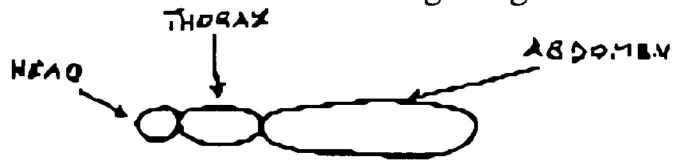
Table 3.8 Overall percentage at each S.O.L.O. response level (from Collis and Davey, 1984, 18).

Item	Percentage of students at each SOLO level			
	Single	Multiple	Relational	General Abstract
G4	51.9	40.7	7.4	0.0
B3	98.5	73.5	20.6	2.9
B4	92.5	66.0	9.4	3.8
B6	80.0	31.7	28.3	15.0
B7	85.7	91.7	53.6	0.0
P3	95.5	58.7	33.3	0.0

An example of a S.O.L.O. item is presented below. The example is item B3, which was used as part of the data gathering S.O.L.O. instrument. The percentages of students classified in the various cognitive levels is given in table 3.8.

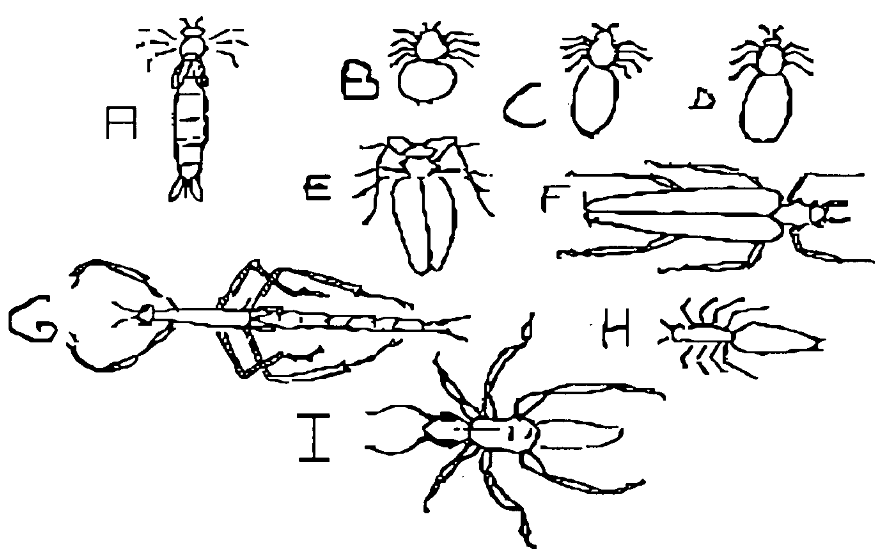
Item B3

Insects have only six legs. Some insects also have wings. Spiders have a fused head with thorax and eight legs.



THE BODY PARTS OF AN INSECT

Below are diagrams of a variety of insects and spiders.



S: What are the three parts of the body of an insect called?

M: Which of the above are spiders?

R: Put all of the above (insects and spiders) into one list in order of increasing body length. Any insects and spiders of equal body length should be grouped together.

G: Devise a dichotomous key or tree diagram for classifying the above insects and spiders.

The format of the S.O.L.O. example above is the same for all other items: a stem with information is presented and four questions follow which the students are expected to answer in order of increasing difficulty. The possibility of misclassification due to response reversals was investigated by Collis and Davey (1986, 657). The probability of misclassification was calculated and found non-significant.

Relating S.O.L.O. levels to Piagetian developmental stages demonstrates clearly that any course that is inquiry-oriented is multimodal in the sense that the student is required to operate at several levels of development. Particularly when hypothesizing, evaluating and generalizing is required the developmental stage is well beyond concrete-symbolic and into formal operations (Biggs and Collis, 1989, 159). The anticipated ages for these levels is well beyond grade ten (Biggs and Collis, 1989, 161). Nevertheless, it has been shown that lower order learning, that is the concrete experiences provided by the inquiry course conducted in this research, facilitates higher order learning (Biggs and Collis, 1989, 158).

In a further investigation into response reversals Wilson (1989, 132) found that especially in Chemistry there was a problem with classifying students into S.O.L.O. levels. The Rasch analysis performed by Wilson (1989, 133-137) showed that all items used in this research were sequential in S.O.L.O. level but that the incremental levels of difficulty were not linear. This meant that in one item the level Relational was less difficult than the Multiple level in another item or that the Relational level in one item may have been more difficult than a General Abstract level in another item. However, on an 80% mastery level of the population as a whole it may be assumed that the S.O.L.O. levels were sequential at the levels Singular, Multiple, Relational and General Abstract. Thus, the SOLO profiles of students on a group basis especially could be accepted with a high degree of confidence making the conclusions drawn in this thesis valid. Therefore, when the responses to the S.O.L.O. instruments were marked, for the items used in this thesis, no adjustments had to be made to allow for the lack of linearity in the levels of difficulty.

3.5 The Student Perceived Characteristics for Success (S.P.C.S.) instrument for students studying science by the inquiry method

3.5.1 Introduction

This section is concerned with the S.P.C.S. instrument, which was designed to meet the need of this research. The aim of the S.P.C.S. instrument was to find out what behaviour characteristics and skills separate the successful from the unsuccessful student using the inquiry method of learning. The terms successful and unsuccessful are defined in the next section.

The reason for the construction of the instrument was how to overcome the dependency on teacher observations in the classification of students as successful and unsuccessful. Any suggestion in the literature invariably used some measure of achievement or ability. Classroom observations, however, seemed to contradict the relationship between success and ability. A number of high ability students were unsuccessful. This instrument establishes criteria other than ability as differentiators between successful and unsuccessful students.

The background to this instrument was that teacher observation seemed to point out that intellectual ability did not seem to be a predominant factor in deciding success. The hypothesis that could be formulated from this was: ability is not a differentiator between successful and unsuccessful inquiry students. There are other criteria which differentiate between successful and unsuccessful inquiry students.

3.5.2 Definitions and characteristics of successful and unsuccessful

Successful means: willing to participate actively in the process of science, that is formulate hypotheses, design experiments, carry out the experiments, look at the results and come to a conclusion. By unsuccessful is meant: not willing to participate actively in the process of science, but instead sit back and get results and conclusions by means other than the inquiry process. It is assumed here that the science method is the inquiry method. Thus, successful students spend more time experimenting productively in the classroom.

Behavioural characteristics of the students were derived by focussing on particular students in several classes over a period of two weeks. A daily logbook was kept to note observations. The characteristics for successful students are recorded in table 3.9.

Table 3.9 Behavioural characteristics of successful students.

Work habits
(1) are not easily distracted;
(2) never interfere with others or only for a quick question;
(3) work in pairs or individually;

Experimentation
(4) happy to fiddle, play;
(5) willing to repeat the experiment or ones similar to it;
(6) willing to try new things;
(7) can suggest alternatives;

(continued over page)

Suggestions and ideas
(8) quick to try out a new suggestion from the teacher;
(9) receptive to ideas from others, including the teacher;

Questions and and answers
(10) want to find the answer themselves;
(11) ask questions about their results;
(12) know what they are trying to achieve;
(13) if shown to be correct tend not to be surprised;
(14) look for answers in their experimental results;
(15) want to show the teacher their results;

Equipment
(16) get their own equipment;
(17) decide what equipment they want before looking for it;
(18) use the equipment they have chosen and watch for results, tend not to ask permission to start the experiment.

The characteristics of unsuccessful students are presented in table 3.10.

Table 3.10 Behavioural characteristics of unsuccessful students.

Work habits
(1) like larger groups more than pairs or singles;
(2) slow to start;
(3) seldom start by themselves;
(4) willing to talk to visitors;
(5) when visiting others may start with question but invariably socialize;
(6) easily distracted;
(7) happy to socialize;

Experimentation
(8) tend to watch rather than participate;
(9) usually watch others get equipment;
(10) are loath to repeat the same or similar experiments;

Suggestions and ideas
(11) seldom suggest new ideas;
(12) seldom suggest alternative experiments;
(13) hesitant to act on a suggestion;
(14) unwilling to do new things;

(15) tend to accept teacher suggestion as truth and loath to try themselves;

Questions and answers
(16) want to be told the answer, unwilling to give their answer;
(17) ask questions on how to start;
(18) tend to be surprised if shown to be correct;
(19) want answers from the teacher not from the experiment;

Equipment
(20) want to show the teacher their equipment selection for approval when they have to get equipment themselves.

3.5.3 Design of the S.P.C.S. instrument

The initial S.P.C.S. instrument contained fourteen items and was administered to four successful and four unsuccessful students. The results of this pilot S.P.C.S. instrument are presented in table 3.11.

Table 3.11 The results of the pilot S.P.C.S. instrument

Item	Unsuccessful students					Item	Successful students				
	SD	D	U	A	SA		SD	D	U	A	SA
1*				2	2	1				1	3
2		1	3			2		3		1	
3*		1		3		3			1	3	
4		1	1	2		4		2		2	
5			2	1	1	5		1	2	1	
6*	1	2	1			6	1	3			
7			1	2	1	7	1			3	
8		1	1		2	8		1	1		2
9*				1	3	9				4	
10*				1	3	10				4	

(continued over page)

11		1		2	1	11	3		1
12*				2	2	12			4
13*				2	2	13			4
14*	2	2				14	4		

Legend: SD= strongly disagree, D= disagree, U= undecided, A=agree, SA= strongly agree.
Note: the numbers in the entries refer to how many students answered in which category. There were four unsuccessful and four successful inquiry students. Those items labelled * showed no significant difference between unsuccessful and successful students.

These fourteen items showed that there was no difference on the following characteristics:

(1) desire for active manipulation. They preferred individual active manipulation;
(2) non-participation in experiments;
(3) ability in interpreting data. They were unable to interpret data;
(4) seeking knowledge or being inquisitive about interpreting data;
(5) seeking confirmation. They equally sought confirmation of results;
(6) preparedness for repetition. They were equally prepared to repeat experiments.

A revised version focussed on those items of the initial fourteen to which successful students had answered very differently than unsuccessful students. The revised version was administered to six successful and six unsuccessful students.

The result of the revised version confirmed that three questions of the original S.P.C.S. instrument were viable items. One reworded question also resulted in separating successful and unsuccessful students.

The revised version showed that successful students felt they were more able to think of alternative ways of doing experiments than unsuccessful students and felt less need to ask the teacher about methodology. Successful students also felt they were better able to plan their experiments and felt overwhelmingly that they need not check with the teacher before doing experiments. Unsuccessful students felt a strong need to check before doing experiments , which showed a difference in perceived confidence in ability to carry out an experiment.

3.5.4 The form of the final S.P.C.S. instrument

The final S.P.C.S. instrument form only investigated four variables:

(1) ability to think of alternatives;
(2) methodology;
(3) planning;
(4) checking with the teacher before doing experiments.

Table 3.12 shows the final S.P.C.S. instrument form.

Table 3.12 Final form of the S.P.C.S. instrument.

This final S.P.C.S. instrument will add to the information you have already given me and will help me in my research.
Read the items carefully and answer truthfully.

1. When I ask the teacher questions they are usually about how to do the experiment.
2. I don't usually decide what equipment and chemicals\materials to use until I can see all the things that have been put out for me to use.
3. I usually check with the teacher before I do the experiment.
4. When I have done an experiment and I did not get any results sometimes I can think of another way to do the experiment and sometimes I can't.

Answer sheet
Place a * on the line to show how much you agree or disagree with the statement;
SA = strongly agree
A = agree
D = disagree
SD = strongly disagree
(You can place a * anywhere on the line)

1.
D________U_______
2.
________ _______ ________SA
3.
________ _______ ________
4.
________ _______ ________

SD_______
A________ SA
SD_______

SD_______
SA
SD_______
SA
SD D
U A
SA

The table 3.13 below shows how to interpret student responses to each of the items.

Table 3.13 Item topic and interpretation of responses to the four items in the S.P.C.S. instrument.

Item	item topic	interpretation of response to item
1	methodology	SA and A: needs help with methodology SD and D: does not need help
2	planning of experiment	SA and A: does not plan ahead SD and D: can plan ahead
3	confidence in carrying out/doing experiments	SA and A: not confident in experimenting SD and D: confident doing experiments
4	think of alternative ways of doing experiments	SA and A: can think of alternatives SD and D: cannot think of alternatives

The final S.P.C.S. instrument was administered to 110 students in five different classes covering a wide range of abilities. At the time the S.P.C.S.

instrument was administered all students had been taught science using the inquiry method for five and a half weeks.

3.5.5 Results of the S.P.C.S. instrument

The data collected in the final S.P.C.S. instrument are presented in table 3.14. To compare the data the responses were calculated on a class average basis. Students were classified successful or unsuccessful by observation, but 26 out of 110 students were unable to be classified as either successful or unsuccessful. This was due to inconsistent behaviour of those 26 students.

Table 3.14 Average scores of all successful and all unsuccessful students on each S.P.C.S. item.

Item	successful	unsuccessful	difference
1	-0.2	1.48	1.68
2	-0.15	1.52	1.67
3	0.04	1.07	1.03
4	1.65	1.90	0.25
no sts	55	29	

Note: SD=-4, D=-2, Undecided=0, A=2, SA=4

Legend: no sts=number of students

The results were further analysed to check if there was a difference in the way low and high ability students responded to the items. In particular, this analysis compared the difference between successful and unsuccessful students of low ability with the difference between successful and unsuccessful high ability students. From the 110 students who participated in the S.P.C.S. instrument there were 29 low ability students, both successful and unsuccessful, and 36 high ability students, again both successful and unsuccessful. The students were classed as high ability and low ability based on the students' responses to the S.O.L.O. items.

The results of this analysis are displayed in tables 3.15 and 3.16. In both tables, the numbers refer to the responses given by successful and unsuccessful students as a group. For instance, consider the low ability students in the class: Green Machine (table 3.15).There were four successful Green Machine students; as a group these four students responded with a total of -2 to item one. This is an average response of -2/4=-0.5 (relatively undecided). To item three these students answered 6 as a group or +1.5 on average. That is, they tended to agree with the statement in item three (not confident in carrying out experiments).

Table 3.15 Results of the final S.P.C.S. instrument for S.O.L.O. Singular and Multiple (low ability) students.

Class	item 1		item 2		item 3		item 4	
	S	U	S	U	S	U	S	U
Green Machine								
class total	-2	2	5	4	6	6	6	6
number sts	4	3	4	3	4	3	4	3
Bubbles & Bangs								
class total	-10	20	-18	22	-14	9	16	20
number sts	13	9	13	9	13	9	13	9
Combined								
total score	-12	22	-13	26	-8	15	22	26
total number sts	17	12	17	12	17	12	17	12

Legend: S=successful
U=unsuccessful
number sts=number of students.

Table 3.16 Results of the final S.P.C.S. instrument for S.O.L.O. Related and General abstract (high ability) students.

Class	item 1		item 2		item 3		item 4	
	S	U	S	U	S	U	S	U
Advanced Science								
class total	2	1	-1	12	-1	0	11	9
number sts	9	6	9	6	9	6	9	6
Chem in the Market place								
class total	18	10	18	8	10	10	22	8
number sts	6	4	6	4	6	4	6	4
Green Machine								
class total	-9	2	-12	0	1	6	9	4
number sts	8	3	8	3	8	3	8	3
Combined								
Total score	11	13	5	20	10	16	42	21
total number sts	23	13	23	13	23	13	23	13

3.5.6 Analysis of the S.P.C.S. instrument results

The statistics used to analyse the S.P.C.S. instrument results followed the procedure suggested by Cohen (1976, 332). A significant difference ($p<0.05$) between successful and unsuccessful students' responses was found. A comparison between low and high ability students was also made. The difference between successful and unsuccessful students of low ability was significantly different ($p<0.05$) to the difference between successful and unsuccessful students of high ability. This significant difference between low and high ability students occurred on all four items.

Item one showed the greatest difference between low and high ability students. This difference diminished progressively from item two to item four. The analysis of students answers showed that item four (thinking of alternative ways of doing an experiment) was not as good a differentiator between successful and unsuccessful students as items one, two and three. For this reason, item four was not included in the conclusions. Therefore, students who answered Strongly Agree to items 1, 2 and 3 would most likely have been unsuccessful students.

The analysis of the responses to the S.P.C.S. instrument revealed four statistically significant differences between successful and unsuccessful students:

(1) unsuccessful students tended to be significantly less confident in experimental methodology, as was shown by the responses to item one;
(2) responses to item two indicated that to a significant extent unsuccessful students tended not to plan their experiments in as much detail as successful students;
(3) responses to item three showed unsuccessful students tended to be significantly less confident in carrying out experiments than successful students;
(4) low ability unsuccessful students had significantly lower perceived planning and experimental methodology ability than low ability successful students. For high ability students this difference between successful and unsuccessful students was insignificant.

The statistical means by which these conclusions were derived were suggested by Cohen (1976, 332). Students responded to items along a sliding scale. The responses were then approximated to the nearest Strongly Disagree, Disagree, Undecided, Agree or Strongly Agree point. It could not be argued from a mathematical point of view that a response of Strongly Agree was twice the numerical value of the response Agree. However, by using the sliding scale responses the data could be treated as interval data, rather than as ratio data. Treating the responses as interval data and taking into account the number of successful and unsuccessful students in each class, averages were used to calculate F ratios. To set up the F ratios the averages of the four items of the successful group were compared with the averages of the unsuccessful group using mean sum of squares values. The comparison between low and high ability students was done in a similar way.

The difference between successful and unsuccessful students was significant ($p<0.05$). Furthermore, the correlation ($r=0.73$) between the averages of successful students and the averages of unsuccessful students was also significant ($p<0.001$) (Cohen, 1976, 332). This indicated that the differences between successful and unsuccessful students were relatively constant on all four items. The F ratio between successful\unsuccessful low ability students versus successful/unsuccessful high ability students was also significant ($p<0.05$). The correlation between the averages of successful/unsuccessful low ability students versus the averages of successful/unsuccessful high ability students was 0.78 and also significant ($p<0.08$).

3.5.7 Validity and reliability considerations

The validity of the S.P.C.S. instrument appeared acceptable. From a philosophical point of view the four areas explored in the instrument touched on the heart of the inquiry method: planning experiments, carrying out experiments, scientific methodology, thinking up alternatives when current methods do not produce results.
Thus, all four items related to the inquiry method. The pilot studies showed that the items measured those inquiry characteristics on which successful and unsuccessful students differed. Feedback from students after the pilot instrument and the revision instrument were administered ensured that students understood the questions.

The Cronbach Alpha reliability coefficient for the S.P.C.S. instrument was calculated for students who were classified as either coping successfully or unsuccessfully with the inquiry method based on teacher observations. Students who could not be classified were not included because the regression equations were built around only those students who could be classified.

The Cronbach Alpha calculations showed that the variance of student scores was much greater than the variance of the instrument items resulting in a very reliable coefficient of 0.93 for successful and unsuccessful students combined. A separate coefficient for unsuccessful students only was 0.82. The latter coefficient was calculated because the analysis of the S.P.C.S. instrument showed that the instrument is particularly suitable for identifying students unable to cope with the inquiry method (if all four items are answered: agree or strongly agree, the student is most likely an unsuccessful student, $p < 0.05$).

To improve the reliability for future research, the number of items could be increased to sixteen by quadrupling each item. The use of negatively phrased items to achieve opposite scores could also improve the reliability.

The S.P.C.S. instrument has proved that the classification successful/unsuccessful based on teacher observations was indeed justified and correct. Probably, the most effective use of the S.P.C.S. instrument is for the identification of unsuccessful students.

3.5.8 Conclusions drawn from the analysis of the S.P.C.S. instrument's results

The single most important result from the S.P.C.S. instrument was that the distinction between successful and unsuccessful students, which was originally based purely on teacher observation, was measurable. The S.P.C.S. instrument provided a basis for quantitatively measuring that successful and unsuccessful students differed significantly in three areas. The fourth area (thinking up alternatives) was not as good a differentiator as the other three:

(1) methodology;
(2) planning ability;
(3) carrying out experiments.

It should be noted that each of these areas separately did not necessarily provide a distinction between successful and unsuccessful. The three areas taken together, with questions 1, 2 and 3 all answered Agree to Strongly Agree or the average of the three approaching Agree (2 points), did provide a significant indicator for unsuccessful students.

To conclude the discussion on the S.P.C.S. instrument it is interesting to relate the S.P.C.S. instrument results to the inquiry method taught to the students. The inquiry method of teaching science is process related, that is the answers/conclusions found are less important than the method by which they are derived. It is, therefore, not surprising that successful students in inquiry classes are separated from unsuccessful students by their experimental methodological ability, even if this ability is only perceived and not necessarily real. The 'bent' toward science that successful students are assumed to have does not appear to be related to ability nor to interest. Perhaps the 'bent' is more related to the perceived confidence these students have in being able to experiment, that is in their perceived confidence in being able to obtain data by experimentation.

The literature review revealed how to teach the inquiry method. This chapter detailed what sudent characteristics were gathered in order to find out not only what differentiates successful from unsuccessful students who were taught by the inquiry method, but also how students taught by the inquiry method compare with students taught in the traditional way. The next chapter will show the details of the framework in which students taught by the inquiry method are compared on the bases of age, ability and branch within science and how they are compared with students taught in the traditional way.

CHAPTER 4

RESEARCH DESIGN

4.1 Research hypotheses

The research hypotheses are stated in two parts. The three statements listed under hypothesis 1 relate to comparisons of students all studying the inquiry method of teaching science. In the second part, the two statements listed under hypothesis 2 relate to comparisons between students learning science by two different methods. The two different methods are the inquiry and the conventional science teaching methods.

4.1.1 Hypothesis 1 statements

Hypothesis 1(a): in the inquiry method of teaching science successful students have significantly different characteristics than unsuccessful students. The student characteristics will be selected from categories of the four instruments described in chapter three (page 40). The selection will be made on a statistical basis taking into account inter-category correlations and correlations to student end of quarter achievement scores.

Hypothesis 1(b): the difference between successful and unsuccessful students is not the same for S.O.L.O. Singular and Multiple students as for S.O.L.O. Relational and General Abstract students. The difference will be measured in terms of the characteristics specified in hypothesis 1(a).

Hypothesis 1(c): students studying Chemistry, Physics and Biology by the inquiry method have significantly different characteristics for each of the disciplines. The characteristics will be those specified in hypothesis 1(a).

4.1.2 Hypothesis 1 details

Hypothesis 1(a) is concerned with the lack of success of the inquiry method of teaching science. The hypothesis aims to quantify the difference between success and lack of success. This is done by determining which student characteristics must be developed or concentrated on to promote the chances of success of the students with the inquiry method. To aid the visualization of the successful student characteristics, a model could show relevant characteristics and relate these characteristics to cognitive development where appropriate.

Due to the nature of the statistics used in the comparison of the five treatment classes it is not possible to develop regression equations to determine the student characteristics needed for success in the inquiry method, because regression equations are developed from parametric statistics. To overcome this barrier success characteristics were determined by using non parametric statistical calculations.

Hypothesis 1(b) addresses the ability variable of success characteristics. It investigates if the inquiry method is more beneficial to S.O.L.O. Relational and General Abstract ability students than to S.O.L.O. Singular and Multiple ability students, as has been claimed in the literature (see page 36). Hypothesis 1(b) is also investigated by the S.P.C.S. instrument developed as part of this research.

Hypothesis 1(c) investigates the claims made in the literature that success characteristics vary not only from subject to subject but also from subject to subject within a discipline (see page 53).

4.1.3 Hypothesis 2 statements

Hypothesis 2(a): the inquiry method produces end of quarter achievement scores equal to or better than the conventional science teaching method.

Hypothesis 2(b): regression equations for the inquiry and conventional teaching methods will be significantly different. The student characteristics to be used in the regression equations will be those determined in hypothesis 1.

4.1.4 Hypothesis 2 details

The difference between hypothesis 1 and hypothesis 2 is that hypothesis 1 uses data that apply only to the treatment classes whereas hypothesis 2 uses matched experimental and control classes. The two hypotheses purposely interweave to provide backup for each other and to allow the fullest understanding of, firstly, how inquiry differs from conventional teaching methods and, secondly, what the characteristics of the inquiry method itself are. Hypothesis 2(a) uses a different methodology than hypothesis 1 in that hypothesis 1 uses all classes involved in the research, whereas hypothesis 2(a) uses only matched experimental and control classes.

With regard to hypothesis 2(a) many controlled experiments have been conducted comparing different teaching methods. The conclusions of those experiments have not been consistent (refer results of research studies, in the Literature Review chapter, page 39). Few experiments have been conducted with the inquiry method as treatment, compared with the discovery and laboratory methods. The main reason for including this hypothesis is concerned with whether the inquiry method is desirable. If more students can be successful in the inquiry method and produce equal or better end of quarter achievement scores than conventional teaching methods, than the inquiry method is worthwhile and desirable. Also, if the inquiry method produces achievement scores equal to or better than conventional methods then at least it is worthwhile from a parent/student point of view. The inquiry method is also worthwhile from educational perspectives (sound from a learning point of view, similarity with actual science process) (Schwab, 1966; Bassler and Kolb, 1971, 112; Fensham, 1981, 55; Driver, 1983, 51, 60 and 86; Rotheram, 1984 and 1987). The development of regression equations in hypothesis 2(b) is essential to be able to determine which student characteristics must be developed to increase the likelihood of success with the inquiry method. This hypothesis ties in closely with hypothesis 1(a).

4.2 Number of participating classes

The research was conducted at Queechy High school. The school is situated in the small city of Launceston in Northern Tasmania. The students come from a city environment. The socio-economic environment of the students falls into two broad categories. One category is the middle to upper-middle class suburbs. These students tend to assume they will continue into further education of some form. The other category is working class. Students from this category typically seek unskilled labouring employment after grade ten, or are oriented toward a trade.

The treatment, to which some of the students were subjected is described in detail in the next section: **4.4 Bases of the inquiry method** (see page 88).

In total six classes participated. There were five treatment classes and one control class. Table 4.1 shows an overview of the participating classes.

Table 4.1 Details of the six participating classes.

Type of participation	grade	Science discipline	topic name
Control	9	Chemistry	Essential Chemistry
Treatment	9	Chemistry	Essential Chemistry
Treatment	9	Chemistry	Bubbles and Bangs
Treatment	10	Physics	Electricity and magnetism
Treatment	10	Chemistry	Chemistry in the Market place
Treatment	10	Biology	The Green Machine

The five treatment classes were taught science by the inquiry method during quarter three, 1990, for twelve week units: three grade tens, two grade nines. The classes did not all study the same science topic, as shown in table 4.1. Teaching all five treatment classes by the inquiry method during quarter three, 1990, allowed the inquiry method to be taught consistently without the need for changing from the inquiry to the traditional style. A grade nine control class, matched to the grade nine chemistry treatment class, was taught during quarter four, 1990, also for twelve weeks. The match was based on student population (grade nine), science discipline (chemistry), science topic (Essential Chemistry) and means of measuring achievement (T.C.E. criteria). The match made the use of parametric statistical analysis possible. All treatment and control classes were taught by the same teacher. This eliminated the teacher bias variable.

4.3 The types of calculations that could be performed to test the hypotheses of this thesis

Statistical analyses, like t, F and Chi square calculations, compare 'groups with groups' (Cohen, 1976, 332). They are known as non-parametric statistics and compare means of groups rather than individuals with individuals (Smith, 1970, 113). Comparing group-means limits the number of degrees of freedom. For instance, two groups with a total of 60 students has a degree of freedom between groups of df= k-1= 2-1= 1 and the degrees of freedom within the groups are df= k(n-1)= 2(60-1)= 118. These two df values are used to calculate the significance of t and F values. The lower the df values the greater the t and F scores need to be for them to be significant (Smith, 1970, 116; Roscoe, 1975, 429).

The degree of freedom, in this study, between the group-means of the five treatment classes was favourable: df= k-1= 5-1= 4. The total number of students in these five classes also resulted in favourable degrees of freedom within the groups: df= k(n-1)= 5(110-1)= 545. Consequently, t, F and Chi square values had a greater likelihood of being significant than a design with only one treatment group. This was one of the main reasons for using non-parametric statistics to analyze the five treatment classes. The other main reason for using non-parametric statistics was that each of the five treatment classes separately was not a random sample distribution of the student population as a whole. The conditions that caused this lack of randomness were: there was no random selection regarding which classes were to be taught by the inquiry method, classes were streamed into high/average and average/low ability, students were from both grade nine and ten.

In contrast, the two matched control and treatment classes were random sample distributions and representative of the high/average ability grade nine student population. Therefore, these matched classes could be analyzed using parametric statistics. The latter techniques were also used by Cheong (1971) in a Ph.D. thesis titled: "An analysis of inquiry performances of High school Biology students including the relationship of inquiry performance to instructional techniques and to student achievement and academic ability".

The comparisons made in the statistical analyses are outlined here to provide a brief overview of the evaluation of the hypotheses. Details are provided in chapter 5.

4.3.1 Comparisons among the five treatment classes

Non-parametric analyses were used to investigate hypotheses 1(a), 1(b) and 1(c). Hypothesis 1(a) compared successful inquiry students with unsuccessful inquiry students and a smaller body of data might have sufficed. Hypothesis 1(b) not only split the body of data up into successful and unsuccessful students, it further divided these two parts in approximately half again in order to compare S.O.L.O. Singular and Multiple students with S.O.L.O. Relational and General Abstract students. Hence, hypothesis 1(b) required a larger body of data than hypothesis 1(a) in order to have reasonable degrees of freedom for significance testing. The five treatment classes only contained one Biology class and one Physics class. This was judged a sufficient number of students to be able to investigate hypothesis 1(c) which was concerned with the disciplines of science.

The following comparisons were made among the five treatment classes to study hypotheses 1(a), 1(b) and 1(c). For each of these classes separately and for the five classes as a whole successful students could be compared with unsuccessful students based on the following characteristics:

(1) achievement scores;
(2) levels in the responses to the Structure of Observed Learning Outcome instrument;
(3) Combined Cognitive Preferences Inventory scores;
(4) scores in the Test Of Science Related Attitudes instrument;
(5) scores in the Cognitive Style of Categorisation Behaviour instrument.

Using these characteristics the following relationships could be investigated:

(1) the separation of unsuccessful and successful students on the basis of both end of quarter achievement scores and the categories of the instruments used to measure student characteristics. These relationships relate to hypotheses 1(a), 1(b) and 1(c);

(2) achievement scores versus each instrument category. These relationships formed the basis of selection of student characteristics in the investigation of hypothesis 2(b);

(3) the relative importance of each instrument category on successful/unsuccessful students. This is a refinement of the relationships investigated in statement (1);

(4) the relative importance of each instrument category on predicting achievement scores but using above and below one standard deviation of the mean category score rather than above and below average category scores. This is a refinement on statement (2) with the aim of extracting clearer relationships;

(5) calculate the correlation coefficients among the different instruments. These relationships would help select the instrument categories to be used in the investigation of hypothesis 1(a) as well as those hypotheses dependent on hypothesis 1(a), such as hypothesis 2(b).

4.3.2 Comparisons between control and matched treatment classes

The data from the matched treatment and control classes were analyzed by means of a parametric analysis. The hypotheses 2(a) and 2(b) compared the inquiry with the conventional method of teaching science. Although two classes may seem small for a statistical analysis, it must be remembered that the parametric analysis compared students on an individual basis. Direct comparisons between compatible treatment and control classes, which studied the same topic and had the same objectives, could be made in two ways. Firstly, comparisons were possible on the bases of achievement scores and each of the instrument categories. Secondly, regression equations for treatment and control classes could be calculated. These regression equations would show the relative importance of each of the instrument categories using the achievement scores as criterion and selected instrument categories, that is student characteristics, as co-variables.

The actual outline of the statistical procedures followed is presented in chapter 5.

4.4 Timing of the administration of the instruments

The instruments used to measure student characteristics are described in chapter 3. The timing of the administrations of the instruments was planned for minimum interference among the instruments to increase reliability with regard to inter-instrument interference, student fatigue and boredom. If student fatigue and boredom occurs the responses might be made in a random fashion rather than answering truthfully.

The S.O.L.O. instrument, Cognitive Style of Categorization Behaviour instrument and Combined Cognitive Preference Inventory instrument are relatively stable over the time period of this research (Scott, 1973, 323). Therefore, it was not critical when they were administered. Hence, they could be divided throughout the quarter with the exclusion of the last few weeks when the students were subjected to many tests in other subjects.

The stability of the T.O.S.R.A. instrument was not known. Therefore, this instrument was administered both at the start and at the completion of the quarter. The test-retest reliability of the T.O.S.R.A. test was reported to be good at 0.78 (Fraser, 1978, 512) and it was considered a valid option to re-administer this instrument.

The S.P.C.S. instrument was conducted about nine weeks into the quarter.

4.5 Bases of collection of achievement scores

For grade nine, student achievement scores on criteria specified for the units of work were recorded at the completion of experiments and tasks. This process went on throughout the quarter. For grade ten, the assessment was not based on particular criteria. Instead, a standard procedure was used; assessing assignments, bookwork, experiment reports and test results. The bulk of the achievement scores was made up of two to three assignments, bookwork at the end of the quarter, experiment reports throughout the quarter and two tests, one mid-unit and one at the end of the unit.

In grade nine, students were marked on criteria according to the Tasmanian Certificate of Education. Usually, the teacher introduced the topic of the experiments and suggested the appropriate criteria for those experiments. A mark of one to eight was awarded depending on the quality of the report as determined by the criteria specified for the unit. If students felt they could improve on a particular criterion the student was invited to either resubmit a better report or satisfy the criterion on another experiment. Thus, at times, one experiment could be used for as many as three or four criteria. Students were made familiar with this system during quarters one and two. This system of obtaining achievement scores was not new to the students in quarters three and four, during which treatment and control classes were run, respectively.

In grade ten, students were awarded marks by taking into account report presentation, accuracy of results obtained, the way the experiment was conducted, detail provided, quality and variety of communication and use of technical language. The students were used to this method since grade seven.

The categories of successful and unsuccessful students were drawn up at the end of the quarter and took into consideration the students' daily behaviour, bookwork, experiment work, general efficiency and effectiveness of time-use in the laboratory.

4.6 Bases of the inquiry method

4.6.1 Philosophy

The philosophy of this research study was to make sure that the inquiry method used here was consistent with the views of the researchers mentioned in the Literature Review chapter. To ease the transition for students, who had never been exposed to the inquiry method, it was decided to start the unit with closed-ended tasks and gradually move to more open-ended tasks toward achieving free inquiry (see page 33).

4.6.2 Aims

The aims of the inquiry method, which were adopted for implementation in the research are listed in eight points. These eight points were the principal issues that guided the formulation of the teaching plans (refer Appendix F).

The inquiry method must:

(1) focus on the process used to reach a conclusion (Schwab, 1966, 67). This is necessary to distinguish it from the discovery method. The latter is end-product (conclusion) oriented;
(2) ensure that content (theory) is clearly presented in context of method (experiment) in order to prevent the artificial separation of content and method (Schwab, 1966, 102);
(3) encourage higher level thinking and prevent minimal responses by students providing instant and correct answers (Wollman and Lawson, 1977, 68);
(4) encourage systematic experimentation by moving progressively from trial and error approaches to experimentation (refer Literature review chapter page 30);
(5) provide students with a realistic image of science by requiring

students to conduct inquiry into science under the guidance of and with the teacher (Tamir, 1985, 93);
(6) allow students to experience both stable and fluid inquiry (Duschl, 1986, 30) by varying the emphasis of the inquiry from closed-ended to open-ended and introduce discrepant events as proposed by Charles (1976);
(7) encourage hunch generation (Wilson, 1974, 131);
(8) make the students accept that doubt is a natural part of scientific inquiry by allowing students to become familiar with a feeling of doubt in the cause of the inquiry (Schwab, 1966, 61).

4.6.3 Strategy for helping students

At the start of the units it was anticipated many requests for help and questions would come from the students. To ensure consistency (relative to both students in the same unit and to students in other units) and validity (with regard to the inquiry method) the following three guidelines were adopted:

(1) students could be helped with process to the extent that they become familiar with the skills and methods of inquiry. Thus, suggestions and choices could be formulated by the teacher, but the student had to make the choice and decide on the course of action. Even if the teacher could foresee that the student would waste time with a wrong choice or embarked on an irrelevant course of action, the student was responsible for making the choice;
(2) the student could ask the teacher for help with reaching conclusions. This distinguishes the inquiry method from the discovery method. In the discovery method the student is expected to reach conclusions independently from the teacher;
(3) the teacher could correct incorrect conclusions. The student could be given a clear instruction to help reformulate a conclusion because the focus of the inquiry method is on the process and not on the conclusion.

4.6.4 Objectives

The objectives of the units were the ones specified in the unit outlines. It was necessary to use them because several units were prerequisites to other units. Also, the students were judged in a relative way with other classes which were taught the same unit in the past.

4.6.5 Homework

Homework, when given, was restricted to researching facts to help finish or reach conclusions or to develop plans of attack for experiments. Standard exercises involving calculations were not given and drill and practice type exercises were totally avoided.

4.6.6 Assignments and tests

Assignments, completed both in class and at home, took on the form of open-ended questions, discovering trends in data, asking for alternative explanations, questions concerning the process of scientific inquiry

Tests were constructed along lines similar to assignments or referred to actual experiments the students had performed.

4.6.7 Student organisation

Students generally worked in pairs, in threes and in fours to encourage cooperation and promote free exchange of ideas and opinions. For this reason group reports rather than individual reports were accepted when appropriate. Students working alone was discouraged.

4.6.8 Equipment organisation

In the beginning of the unit all necessary equipment was provided for each experiment to help students start. After three weeks sometimes irrelevant material was included with relevant equipment. Later, only some equipment was provided, enough to give students ideas but they needed to supplement equipment themselves by writing equipment lists. In the last few weeks, very little equipment was provided and the students had to decide on equipment lists.

The laboratories which were the venues for the classes all had standard equipment permanently displayed in baskets, for instance glassware, crucibles, retort stands and Bunsen burners.

4.7 Examples of teaching plans for the inquiry method

This section outlines some details of how the treatment classes were taught. These details are extremely important because it is the crux of the inquiry method. One of the criticisms that can be levelled at other research in this area is that not enough details of the actual learning tasks are provided. Lack of detail provides difficulties with replication and judgement of the details of the inquiry method used. In the Literature Review chapter (see page 29) it was pointed out that many interpretations of the inquiry method exist. To facilitate replication and to make explicit the inquiry method used for the treatment groups several examples of teaching plans are provided.

There were five treatment classes. In this section examples of teaching plans are presented for each class. A full list of teaching plans can be found in Appendix F. By teaching plans are meant tasks planned for students to perform. They do not indicate objectives and aims, which were discussed earlier in this chapter. All lessons were planned in such a way that all or most of the eight aims of the inquiry method were fulfilled. The sequence of lesson plans presented the students with a series of related problems which systematically developed increased understanding of relationships and application. With regard to the variables used in the statistical analysis, early in the sequence memory Recall and Application were emphasized but not to the exclusion of Principles and Questioning (C.C.P.I. instrument). Later in the sequence Principles and Questioning were emphasized. No particular effort was made to make science more related to the students' everyday life, although Application was emphasized; students' enjoyment of science was not given special consideration either in the design of the lesson plans.

4.7.1 Chemistry in the Market place (grade ten, students mainly S.O.L.O. Relational and General Abstract)

1. Use one, some or all of the following materials to test if samples are acidic or basic (alkaline).
Material: phenolphtalien indicator (p-ind.), methyl-orange indicator, iron filings, Mg strips, Al strips, pieces of Zn, copper oxide.
Samples: lemon juice, HCl (dil.), vinegar, sulphuric acid, sodium hydroxide, potassium hydroxide, sodium bicarbonate, apple, toothpaste, lime, milk, orange juice.

In the report include in your conclusion how to test for acids and how to test for bases.
Time: 4 periods.

2. Some acids are strong, some are weak. Using a base of given strength suggest a method to compare the strengths of citric acid and vinegar.
Material: p-ind., universal ind., burettes, 0.5 M NaOH, vinegar, solid citric acid.
Time: 3 periods.

3. Plan and carry out experiments to determine if the following compounds

are ionic or covalent.

Equipment: water, power packs with volt- and ammeters.

Material: nitrates, carbonates, chlorates, chlorides, sugar, salt, alcohol, ammonia (dil.), kerosene, petrol, wax, fat, petroleum jelly, lubricating oil, pentane.

Time: 4 periods.

Examples of Assignments

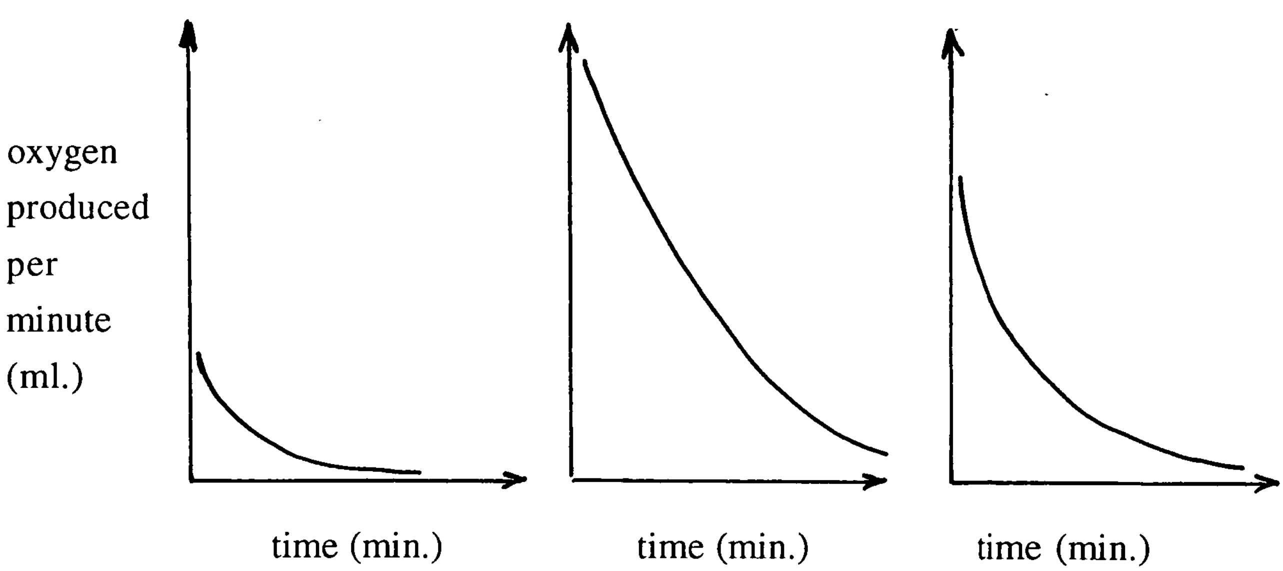
1. Hydrogen peroxide is a liquid with strong bleaching power. It has a limited shelflife, that is it deteriorates quickly. It decomposes into water and oxygen.

The graphs below show how much oxygen is produced:

(a) when hydrogen peroxide decomposes naturally;

(b) when manganese dioxide is placed in the hydrogen peroxide (MnO_2 is a solid);

(c) when gold is placed in the hydrogen peroxide.



Note: all three graphs used the same amount initial volume of hydrogen peroxide.

(i) How do the three graphs differ?

(ii) Can you draw any conclusions?

(iii) What do the three graphs have in common?

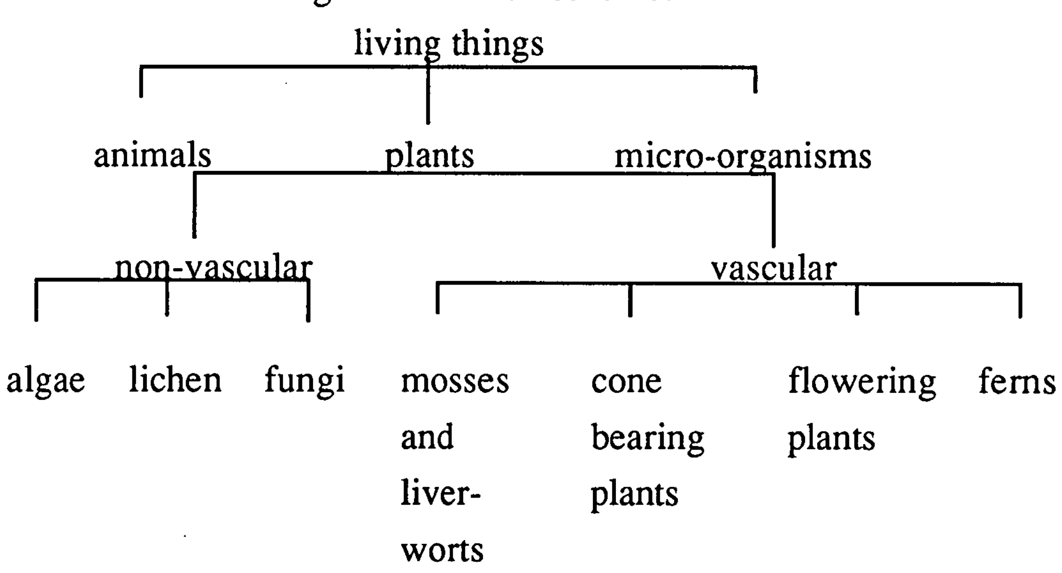
(iv) Does this contradict your conclusion in (ii)? If yes, revise your conclusion.

(v) Look at the table below. Can you draw any conclusions?

Chemical name	mass of chemical before placing in hydrogen peroxide (grams)	mass of chemical after placing in hydrogen peroxide with the reaction completed (mass in grams)
manganese dioxide	5.01	5.01
gold	8.43	8.43

4.7.2 The Green Machine (grade ten, S.O.L.O. Multiple to General Abstract)

1. Given the following classification scheme:



Collect samples from outside and use the specimens provided to determine how you can distinguish between vascular and non-vascular plants.

Material: liquid dyes (2 colours), water, scalpels, wooden boards, microscopes, hand lenses.

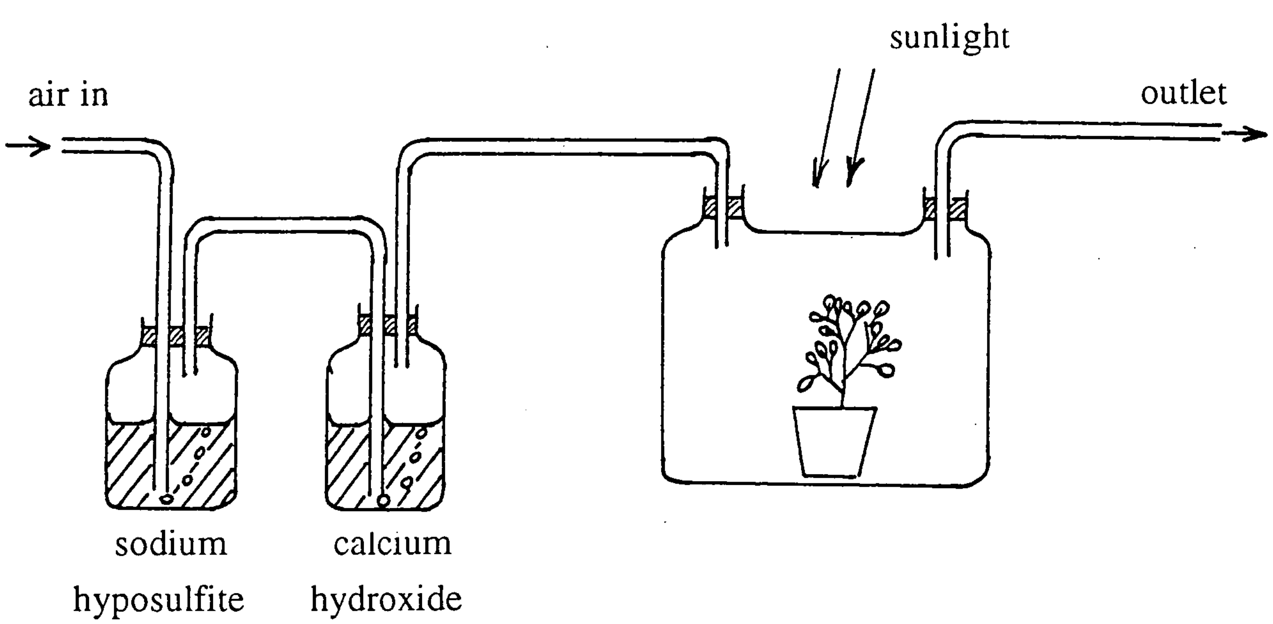
Time: 5-6 periods.

2. Given: 7 commercially prepared microscope slides of vascular and non-vascular plants.
Compare these slides with the characteristics developed in 1. to distinguish between these two types.
Time: 1 period.

3. What gives plants their green colour?
Material: mortar and pestle, metho, filter paper, acetone, iodine, chalk powder, aluminium oxide, glass tubing, cotton wool.
Ask for anything else you want.
Time: 5-6 periods.

<u>Examples of assignments</u>

1. What would the following experiment prove?



Method: An air pump pushes air through the two bottles. The plant was removed from the classroom and placed in the large jar when the air started to flow through.

Results: (i) Iodine confirmed starch is present in the leaves after several hours.
(ii) at the outlet the air was bubbled through Bromothymol-blue and it stayed blue.
(iii) at the outlet the air was tested for oxygen and no oxygen was found.

2. How could the experiment in 6. be improved?

4.7.3 Electricity and magnetism (Grade ten Advanced science, S.O.L.O. Relational and General Abstract)

1. Devise a test for measuring magnetic force quantitatively.
Material: bar magnets, magnetic compasses, string, force balances, wire, power pack, iron nails, iron filings, clear plastic sheets.
Time: 4-5 periods.

2. Use the test devised in 1. to measure if the magnetic force of weak magnets can be improved.
Time: 2 periods.

3. Magnetic field lines are imaginary lines with arrows that indicate the direction in which a compass needle would be forced if the needle were placed on that spot.
(i) Investigate magnetic field lines for a single magnet.
(ii) As (i) for 2 magnets.
(iii) How do strong magnets compare to weak ones?
Time: 0.5 period.

Example of a worksheet on alternative explanations

This worksheet was given to students to demonstrate the need for thoughtful planning during the methodology phase of experimentation. The worksheet illustrates the necessity of controlling all relevant variables.

Below are given experimental data and correct, or partly correct, conclusions. For each suggest an alternative, correct conclusion.

1. To prove that magnetic fields act in 3-D, a scientist filled a small beaker with glycerol (a syrupy liquid more dense than water) and mixed in some iron filings till they were evenly distributed. She then suspended a small magnet in the beaker but observed no change in the iron filings.

Conclusion: magnetic fields do not act in 3-D.

Alternative:

<u>Examples of assignments</u>

1. In separate experiments the following data were collected:

GRAVITATIONAL FIELD			ELECTRIC FIELD		
force on mass (N)	grav. field strength ($N.kg^{-1}$)	mass placed in field (kg)	force on charged particle (N)	electric field strength ($N.C^{-1}$)	charge of particle placed in ($C*10^{-19}$)
12.5	10.0	1.250	$-4.8*10^{-14}$	$3.0*10^{5}$	-1.6021
26.0	130.0	0.200	$+3.8*10^{-15}$	$2.4*10^{4}$	+1.6021
1.0	10.0	0.100	$+3.2*10^{-13}$	$1.0*10^{6}$	+3.2042
1.2	0.1	12.350			

Note: C=Coulomb, N=Newton.

(a) For each of the fields separately can you see any trends?

(b) Derive a mathematical equation for each field.

(c) Can you infer any relationships from these trends?

(d) Are there any similarities between questions 1,2,3 and the data in question 4?

(e) Based on your answers to (b), (c) and (d), hypothesize an equation that would apply to a magnetic field. Identify all variables and explain the relationship.

4.7.4 Bubbles and Bangs (Grade nine, S.O.L.O. Singular and Multiple)

1. Determine the boiling points of the following liquids:
water, alcohol, glycerol, olive oil, hexane (fume cupboard!)
Question: what is a boiling point?
Equipment: Pasteur flasks.
Time: 3 periods.

2. How are the boiling points affected when the liquids are mixed?
Time: 2 periods.

3. From 2. write two questions. Formulate a hypothesis from each question and test one of the hypotheses.
Time: 2 periods.

4.7.5 Essential Chemistry (Grade nine, S.O.L.O. Relational and General Abstract)

1. Do at least two physical changes and two chemical changes. Report on why they are physical or chemical changes.
Material: Fe filings, powdered S, magnet wrapped in tissue, olive oil, salt, methylated spirits, HCl (dil), vinegar, sodium bicarbonate, iodine (solid).
Time: 2 periods.

2. Do experiments to prove that the substance is pure or a mixture. You can use any of the following methods: filtering, evaporating, draining, centrifuging, pipetting, decanting, chromatography, preferential dissolving, looking under the microscope.

Material: salt, sugar, metho, water, milk, orange juice, Fe and S mixture in 1:1 proportion by volume both before heating and after heating (in fume cupboard!), margarine.

3. Which of the elements below are metals? Which are non-metals? Why are they metals or non-metals?
Material: water, acids, bases, battery, ammeter, several leads, universal ind., several metals and non-metals, both to experiment with and to observe only, for instance Hg.
Teacher demonstrations on Na, Li, P, etc., with water and tested afterwards with universal ind.

As always, present your data in several different ways, find patterns and trends, draw conclusions. Then, ask new questions.
Time: 3 periods.

Examples of Assignments

1. 10 g. of blue copper sulphate crystals, a pure substance, was dissolved in water. A positive and a negative electrode were placed in the solution and a low voltage was applied.

An ammeter showed a current flowed. Also, it was noticed that a reddish-orange deposit covered the negative electrode. Nothing was deposited on the positive electrode.

After several days, the reddish-orange deposit was quite thick and the solution was much less blue. The ammeter showed that the current was almost stopped eventhough the same voltage was still applied to the electrodes.

The power was switched off and the negative electrode was weighed: the original electrode mass was 30 g., the final mass of the electrode was 33 g.

(a) How many grams of copper sulphate were dissolved?

(b) was the copper sulphate pure?

(c) What was the reddish-orange deposit?

(d) Where did the deposit come from?

(e) How would you prove that the deposit did not come from the water?

(f) Why did the solution become less blue?

(g) Why was the current nearly stopped after several days?

(h) Why did one of the electrodes weigh more afterwards?

(i) List all the things you know about the copper sulphate solution (I can think of 13).

(j) The copper sulphate was pure, or was it not?
What seems to be the contradiction?
What could this suggest for some pure substances?

4.7.6 Teacher demonstration titled: Unexpected Events

The following teacher demonstration, together with a worksheet, was performed in the three chemistry classes in the second week of the quarter. The three classes were: Chemistry in the Market place, Bubbles and Bangs, Essential Chemistry.

Students were given the following worksheet.

1. Record all observations. Write down your prediction of what will happen before I do the experiment.
2. Look at all your observations and organize the data in several different ways, such as keys, tables, grouping.
3. Can you see a pattern or trends in the data?
4. Write down at least three questions by looking at your data.
5. Formulate hypotheses.
6. Choose one hypothesis. List all the things you can think of that could effect the results (observations) if you were to test the hypothesis.

7. Suggest a method that would determine which of the things you thought of in 6. will effect your results most.
8. Write an equipment list, including quantities, and show me for approval. Some materials are listed at the bottom of this sheet.
9. Do the experiment.
10. Record all your observations, present the results in different ways. That is, start again at point one, until you are able to come up with the same results that occurred in my demonstration.

The teacher demonstration is described next.

Material: several acids and bases, several indicators, nitrates, sulphates, iodides, carbonates.

Display five beakers. The beakers are all filled equally with 200 ml. They all look the same: colourless, clear liquids.
Beaker 1: water and p-ind.
Beaker 2: H_2SO_4 and p-ind.
Beaker 3: a saturated solution of $Pb(NO_3)_2$.
Beaker 4: very dilute solution of $Pb(NO_3)_2$.
Beaker 5: a $MgSO_4$ solution.

To each of these beakers in turn, add slowly 200 ml. of NaOH. The final results are five different coloured solutions.
Time: 2 periods.

4.8 Closing remarks

Although it is partly known why students find the inquiry method difficult it is not known why some students are better able to cope with the inquiry method than others (Schwab, 1966; Seymour, Padberg, Bingham and Koutnik, 1974;

Welch, Klopfer, Aikenhead and Robinson, 1981; Rotheram, 1984). If the characteristics of successful students are known these characteristics can then be developed and promoted to give more students a chance of success. To facilitate communication of relevant characteristics the development of a model would be useful. Students are not programmable computers. Not every student can be expected to succeed because abilities, skills, attitudes, backgrounds and personalities vary. Student characteristics also change, they are 'dynamic, fluid and ever-changing' (Tamir, 1977, 479). Most are, however, stable over periods of one term to one year (Scott, 1973, 323). The determination of characteristics of successful inquiry students became the basis of this thesis and shaped the design of this research.

This chapter provided the details of the experimental design. These details included the research hypotheses, a description of the treatment and control classes and how and what these classes were taught. The next chapter will show exactly how the data collected during this study, using the instruments described in chapter three, will be analyzed statistically in order to find answers to the hypotheses stated in this chapter.

CHAPTER 5

THE STATISTICAL METHODOLOGY: A DESCRIPTION AND RESULTS

5.1 An explanation of the two pronged statistical analysis

The research design used five treatment classes to ensure consistent inquiry teaching (by preventing the teacher having to adjust from inquiry to traditional from class to class), obviate teacher bias between different classes and to be able to compare different student populations (grade nine and ten; high/average and average/low ability; different science topics). However, the drawback of this design is that each class separately is no longer a random sample student population as a whole. Parametric statistics assume that the individuals in the sample classes are drawn randomly from a population which is distributed normally (Cohen, 1976, 316). To eliminate this element of doubt, non-parametric statistics were used to analyze the treatment classes. Non-parametric statistics are less sensitive to violations of normality of distribution and selection of individuals into samples because they use group-means as the basis of comparison (Dayton, 1970, 33). The research design matched a treatment and a control class by drawing students randomly from the same population which could be assumed to be distributed normally (high/average ability grade nine students). This match allowed parametric statistics to be used to compare these two classes.

The mechanics of parametric and non-parametric statistics are vastly different. The former is able to compare students on an individual to individual basis. The latter can only compare on the basis of averages. As a consequence the results of non-parametric statistics are limiting. These limitations were overcome by sequencing a series of steps of non-parametric techniques in which the next step relied on the results of the previous step. By sequencing the steps more detailed trends could be achieved and the conclusions had the potential to be more far reaching. Therefore, the data collected from the six classes (five inquiry classes and one control class) were analysed along two separate paths. One path compared the five inquiry classes using non-parametric statistics. A second path compared the control class with a matched inquiry (treatment) class using parametric statistics.

5.2 The non-parametric analysis

Diagram 5.1 Flowchart of the non-parametric statistics used

Five treatment classes; student characteristics taken from C.S.C.B., C.C.P.I., T.O.S.R.A., S.O.L.O. and S.P.C.S. instruments and achievement scores

U vs S sts by achievement scores
cf: U with S sts on the basis of rank order achievement scores; class by class and all five classes together.
Use: Mann-Whitney U test.

U vs S sts by CCPI and CSCB
cf: U with S sts on the bases of sts grouped in categories of CSCB and

Achievement scores vs CCPI, CSCB and TOSRA
cf: st achievement scores with st scores in three instrument categories.
Use first: Spearman's rho correlation between achievement scores and scores in categories of CCPI, CSCB and TOSRA instruments.

Use second: Kendall's coefficient of Concordance W to determine if the relationships for which rho was calculated

CCPI data; class by class, all five classes together and all sts by SOLO level.

Use first: Chi-square for k samples to establish significant relationships

Use second: Guttman's lambda to quantify the relative strengths of significant relationships

is significant. If insignificant delete those categories with the lowest rho and recalculate W till W is significant.

U vs S sts by CCPI, CSCB, TOSRA and SOLO

cf: U with S sts on the bases of various category combinations of four instruments.
Further refinements include combinations of above and below median scores in categories and correlations between U and S category combinations.

Use: the Weighted Net Percentage Difference (WNPD) analysis of variance to quantify by percentage differences between U and S sts in terms of combinations of instrument categories.
(While Guttman's lambda shows the relative strengths of the U and S differences, the numbers have little meaning in absolute terms. WNPD percentages will be easier to interpret).

Legend: U= unsuccessful inquiry students
S= successful inquiry students
sts= students
cf= compare

5.2.1 An overview of the sequence

The sequence is described in detail in the next section, but an overview of the different techniques used is shown in diagram 5.1. The diagram shows that the Mann-Whitney U test tied together unsuccessful and successful students on the basis of criterion score. The Chi-square test and Guttman's lambda tied together unsuccessful and successful students on the bases of the instrument categories used: Combined Cognitive Preference Inventory and Cognitive Style of Classification Behaviour. The combination of Kendall's W and Spearman's rho tied together the criterion scores and the instrument categories used: C.C.P.I., C.S.C.B. and T.O.S.R.A. In this way, it was felt, a logical connection could be surmised in an algebraic analogy to: if A is related to B and A is related to C, then B must be related to C. This procedure was adopted to show that the triangle: unsuccessful/successful at one apex, the different instruments that were administered at another apex and criterion score/end of quarter achievement score as indicators of student achievement at the third apex, represented real, significant and important relationships.

5.2.2 The non-parametric techniques in detail

The Mann-Whitney U test

For this test the students were categorised as successful or unsuccessful in coping with the inquiry method as determined by teacher observations over a period of one quarter. The correctness of the teacher based categorization was confirmed by the results of the S.P.C.S. instrument (see page 73). Prior to being categorised the students were ranked by their average criteria scores (for grade nine) or by their end of quarter achievement score (for grade ten). The Mann-Whitney U test was used to compare successful with unsuccessful students by rank to determine if there was a significant difference between them. This test was applied to the five classes separately as well as to the five classes combined. The latter comparison required a re-ranking of the subjects on a whole group bases.

The Mann-Whitney U test (Nunnally, 1975, 296) is a non-parametric alternative to a one-way analysis of variance. The test applies only to a two sample situation. For a k-sample situation the Kruskal-Wallis test should be used. Although a continuous distribution is required, the test requires neither normality of distribution nor homogeneity of variance and is almost equivalent to the F test in parametric statistics (Roscoe, 1975, 304). The test does require data at a minimum of ordinal level.

With the students ranked in the sample and then categorised as successful or unsuccessful, the total rank score of each category can be calculated by adding the ranks in each category. The total rank scores are converted into U scores and then into z scores. The z scores are then tested for significance in representative areas under the normal curve (Roscoe, 1975, 425).

The Chi-square test for more than two samples

The Chi-square test can analyse data at nominal level, that is, data without a numerical value but represented as categories. The student categories employed in this research were unsuccessful and successful.

The variables on which the students were compared were derived from the instruments Combined Cognitive Preference Inventory and Cognitive Style of Categorisation Behaviour. In order to achieve a sufficient number of data all students in the five classes were combined. The Chi-square test, therefore, was used to determine if successful and unsuccessful students were significantly different on the variables Combined Cognitive Preference Index and Cognitive Style of Categorisation Behaviour.

The Chi-square test compared the observed frequencies of different instrument categories for successful, unsuccessful and unclassified students with the expected frequencies. The expected frequencies were evenly distributed over all categories in accordance with a randomly distributed population. That is, one would have expected equal numbers of students in each instrument category. At first, it was decided that to be categorized in a particular category the student's score in the instrument for that category had to be one standard deviation above the mean for that category. Later, in order to increase the number of students in the categories, the Chi-square test was repeated using students with above the mean scores rather than one standard deviation above the mean scores.

To perform the Chi-square test a table was constructed with the categories successful, unclassified and unsuccessful as columns. The rows of the table were made up of the categories containing the instrument variable being examined. For instance, in the case of data from the Combined Cognitive Preference Inventory instrument the rows were made up of the categories: Principles, Questioning, Recall, Application, Questioning-Recall and

Principles-Application. The total number of students in a particular element in were then entered the table . The table below shows the actual student numbers in each if the Combined Cognitive Preference Inventory categories. These data were compared with expected frequencies in the Chi-square test.

Table 5.1 The number of unsuccessful, unclassified and successful students who scored one standard deviation above the mean of the particular C.C.P.I. instrument categories.

	Student score one standard deviation above the mean		
C.C.P.I. category	unsuccessful	unclassified	successful
Relational	5	5	7
Principles	4	5	10
Questioning	5	5	11
Application	3	7	10
Questioning-Relational	3	7	7
Principles-Application	5	7	9
Total	25	36	54

Table 5.2 The number of unsuccessful and successful students whose scores were above the mean of the particular C.C.P.I. instrument categories.

	Student score above the mean	
C.C.P.I. category	unsuccessful	successful
Relational	15	29
Principles	14	25
Questioning	11	29
Application	6	25
Questioning-Relational	10	24
Principles-Application	16	30
Total	72	162

In order to obtain a similar table for expected frequencies it was suggested in the literature (Cohen, 1976, 349) that equal numbers of successful and unsuccessful students ought to be expected, that is an average frequency. This average frequency was calculated by adding up all successful and unsuccessful students and dividing by the number of rows available. Similar expected frequencies were calculated for one standard deviation above the mean and above the mean scores.

The disadvantage of the Chi-square test is that no conclusion can be made regarding the strength or weakness of the relationship if one exists. To overcome this limitation a correlation coefficient can be calculated to provide an expression for the strength of the relationship once a relationship has been shown to exist.

<u>Guttman's Coefficient of Predictability, lambda</u>

The Phi coefficient expresses the strength of the relationship for 2x2 tables. The Guttman Coefficient of Predictability is the appropriate correlation coefficient for tables other than 2x2. One of the tables constructed in this statistical analysis was a 3x6 in the case of successful, unclassified and unsuccessful categories as columns and the six C.C.P.I. categories: Principles, Questioning, Recall, Application, Questioning-Recall, Principles-Application as the rows in the table. Another table was also a 3x6 table with, again, successful, unclassified and unsuccessful categories as columns and the six Cognitive Style of Classification Behaviour: Part-whole, Whole, Functional-Interdependence, Categorical-Functional, Categorical-Class, Categorical-Inferential, as the rows in the table.

In the Guttman table each cell contains the number of students, called frequency, that fall into a particular category. Lambda gives an indication of how much a particular frequency stands out. The formula used to calculate lambda is given by Cohen (1976, 352).

There is no test for significance available for lambda. Therefore, it is important to perform a Chi-square test first to determine if there is a significant relationship and only then can lambda be calculated and have meaning. Thus, lambda on its own has little meaning. It simply states that the greater lambda the stronger the relationship for whatever association there exists between the columns based on the row categories used in the determination of lambda. A lambda of zero means there is no relationship at least as far as highest frequencies trends are concerned. Cohen (1976, 353) used examples in which lambda varies from 0.1 to 0.3. In this research the values varied from 0.03 to 0.17.

Because there is no significance test available the absolute size of lambda is less important than the relative size of lambda.

The results of the Guttman Coefficient of Predictability calculations were used to determine which instrument categories were the strongest differentiators between successful and unsuccessful students. These differentiators were later used in the construction of regression equations. The Guttman coefficient also helped to determine whether above the mean scores or one standard deviation above the mean scores were to be used to differentiate between successful and unsuccessful students. Guttman's lambda also provided clear data about the difference between successful and unsuccessful students at different levels of cognitive development, that is at different S.O.L.O. levels.

Spearman's rank order correlation coefficient and Kendall's coefficient of concordance

These two coefficients are totally separate statistical calculations. They do not have to be used together. In the analysis of the data in this research it was considered useful to use these two coefficients in tandem (Cohen, 1976, 371). Spearman's rank order correlation coefficient, rho, was used to calculate the correlation between each instrument category and the end of quarter achievement score. The correlations between each possible pair of instrument categories were also calculated. The correlation data thus obtained were used to eliminate those instrument categories which were least correlated with the criterion score. By least correlated are meant those correlations close to zero;

high negative correlations were still included because there clearly was a relationship even if it was inversely related. High correlation data were values greater than 0.3 absolute.

Kendall's coefficient of concordance, W, was calculated next. The students of treatment classes, one class at the time, were ranked according to their percentage criterion score; this was the score the student was awarded at the end of the quarter. Then, the student's scores on each of the instrument categories separately were entered in a table next to the student's criterion score rank. The instrument categories used were: Combined Cognitive Preference Inventory, Cognitive Style of Classification Behaviour and Test Of Science Related Attitudes. W incorporates four aspects (Cohen, 1976, 367):

(1) the sum of the squared individual deviations from the class mean;
(2) the number of students;
(3) the number of instrument categories;
(4) a correction for tied ranks.

Thus, W provided an indication of how closely the different instrument categories are related with the criterion score. This relation was restricted to a rank basis for both criterion score and instrument score, that is the data were on an ordinal scale. The greater the value of W the closer the agreement between the different instrument categories and the criterion scores. W can be tested for significance using the Chi-square table (Cohen, 1976, 367).

The method of combining W and rho was suggested by Cohen (1976, 371). If W was non-significant in the first calculation, instrument categories with low, absolute, correlation values were eliminated and W was re-calculated. This process was repeated until a significant W resulted. A significant W in turn meant that the correlation data had to be significant as well. Hence, both values, W and rho, could be used to make meaningful conclusions. W and rho data were calculated for all treatment classes.

The Weighted Net Percentage Difference (W.N.P.D.) analysis of variance

The final non-parametric statistical technique used related four variables in one table. The four variables were: the successful/unsuccessful student classification, the instrument categories, student S.O.L.O. levels and end of quarter achievement scores. The W.N.P.D. technique (Cohen, 1976, 379) provided information on a percentage likelihood basis with regard to certain categories of students more likely to belong to one instrument category than to another. For instance, the following information was calculated: a student classified Part-Whole (C.C.P.I.) and Principles (C.S.C.B.) with a high S.O.L.O. rating (e.g. General Abstract) was 63% more likely to succeed in an inquiry science course than a student classified Whole (C.C.P.I.), Application (C.S.C.B.) with a low S.O.L.O. rating (e.g. Multiple).

The W.N.P.D. table was set up as follows. Having previously determined in Guttman's lambda that the strength of the relationship between successful/unsuccessful students and criterion score was stronger for above the mean values than for one standard deviation above the mean, two columns were drawn up. One column contained all students who scored above the criterion mean, the other column contains all students who scored below the criterion mean. Each of these two columns was in turn separated in two columns: one for successful students, the other for unsuccessful students. Each of these four columns was then divided, in turn, in four smaller columns. These final four columns were for the four S.O.L.O. levels used in this research. Therefore, there were, in total, sixteen columns.

To create a matrix from this table each instrument category made up one row which intersected all sixteen columns. Based on the previous statistical calculations (Mann-Whitney U test, Chi-square, Guttman's lambda, Spearman's rho and Kendall's W) only those instrument categories with significant ($p<0.05$ or $p<0.08$) relationships between criterion score and successful/unsuccessful student classification were used. These instrument categories were:

(1) Questioning-Recall (C.C.P.I.) above the mean;
(2) Questioning-Recall (C.C.P.I.) below the mean;
(3) Principles-Application (C.C.P.I.) above the mean;
(4) Principles-Application (C.C.P.I.) below the mean;
(5) Whole (C.S.C.B.) above the mean;
(6) Categorical-Functional (C.S.C.B.) above the mean;
(7) Categorical-Class (C.S.C.B.) above the mean;
(8) Attitude to science (T.O.S.R.A.) above the mean;
(9) Enjoyment of science (T.O.S.R.A.) above the mean.

Therefore, the matrix contained 16x9=145 elements or cells as Cohen (1976, 379) called them.

Four different types of comparisons could be made:

(1) block to block;
(2) row to row;
(3) column to column;
(4) cell to cell.

An example of type one, block to block, compared above the criterion mean score successful to above the criterion mean score unsuccessful students on the basis of all instrument categories. That is, the block of columns one to four was compared to the block of columns five to eight and took into account all the instrument categories.

Row to row comparisons looked at individual instrument categories and could ask questions similar to the following example. How many more times, as a percentage likelihood, was a student classified above average Questioning-

Recall more likely to score above the mean criterion score than below it? This would have compared students only on the basis of above average Questioning-Recall regardless of S.O.L.O. level or whether they were successful or unsuccessful students.

A column to column comparison kept selected instrument categories constant. It could ask: how many more times, as a percentage likelihood, was a student classified successful and S.O.L.O. General Abstract more likely to score above the mean criterion score than below it? This comparison would have taken into account all instrument categories.

A cell to cell comparison was very specific. This type of comparison isolated one particular instrument category and one particular S.O.L.O. level. Comparisons could be made comparing above the criterion mean score successful students to below the criterion mean score successful students or to above the criterion mean score unsuccessful students.

Regardless of which type of comparison was made the method used to calculate the percentage difference is the same (Cohen, 1976, 379). If a block to block comparison was made, then cells in corresponding places in the blocks were compared directly. The algebraic difference of the number of subjects in the two cells was calculated. This difference is multiplied by the total number of students in both cells. It was important to be consistent with the algebraic order in which the cells were subtracted. The order depends on the way in which the comparison was worded. To calculate the W.N.P.D. the individual cells' products are added and divided by the total number of students in all the cells being compared.

There is no method for calculating if the percentage obtained is statistically significant. Hence, it is imperative to perform Spearman's rank order correlation coefficient and Kendall's coefficient of concordance calculations prior to the W.N.P.D. calculations to establish the significance of the instrument categories being included in the W.N.P.D. table.

A total of 55 comparisons were made. The comparisons ranged from general block to block comparisons to specific cell to cell comparisons. General comparisons compared successful with unsuccessful students. Specific comparisons compared for instance unsuccessful students, scoring above average on Attitude but who had below average achievement scores, with unsuccessful students, scoring above average on Questioning-Recall but who had below average achievement scores.

Although no direct calculations could be made to determine if the W.N.P.D. results were significant, a measure of significance could be obtained by correlating one set of cell to cell comparisons with another set of cell to cell comparisons. One set compared, for instance, successful above average (on achievement scores) students with successful below average students. Another set of cell to cell comparisons compared, for instance, unsuccessful above average students to unsuccessful below average students. In this way a correlation coefficient could be calculated and then tested for significance by converting the correlation to a t-score (Cohen, 1976, 332). The t-score was tested for significance at N-1 degrees of freedom, where N was the total number of W.N.D.P. values used in the correlation calculation. The formula to calculate the t-score is given below.

$$t = r \times (S_{r0})^{-1}$$

$$S_{r0} = (N - 1)^{-0.5}$$

t= t-score
r= correlation value
N= number of students

(Cohen, 1976, 332)

This method of determining if the correlation coefficient was significant allowed the overall picture to be developed to gain insight into the differences between successful and unsuccessful students and related these differences to which students scored above or below the mean on achievement scores.

To conclude the outline of the statistical method in the non-parametric part of the statistical analysis, it may be helpful to emphasize that basically only six non-parametric techniques were used. The six techniques were applied to different configurations of the data. It may seem that the profusion of calculations results in an analysis of the same thing. However, as explained in the introduction to this statistical outline there is no one tool available to analyze the data in one calculation. Different techniques have different uses and each technique has a role to play to gain as much information out of the data as possible. It is especially difficult to determine the significance of the results in non-parametric statistics. Here, the symbiotic relationship between the six techniques comes to the fore: the significance of the results of one technique is determined, the next technique builds on this and provides more detail while it relies on the significance of the previous calculation.

From the results of the non-parametric analysis a model was developed to illustrate a sequential development of inquiry student characteristics. This research found that students who coped successfully with the inquiry method had significantly different characteristics than students who were unsuccessful in coping with the inquiry method. Furthermore, the research data suggest that the characteristics of successful inquiry students changed with cognitive development, as measured by the S.O.L.O. instrument. For instance, it was found that students at the cognitive development stage corresponding to S.O.L.O. singular concentrated on part-whole attributes of phenomena and objects in their experiments. This emphasis on attributes changed to a concentration on Principles and Questioning, with little emphasis on Recall of facts, when students reached the S.O.L.O. General Abstract stage of cognitive development. The desirability for the development of this model, showing successful inquiry characteristics, becomes more important in light of the fact that inquiry students performed significantly better than conventionally taught students. The model was developed from the results of the non-parametric analysis and tested and refined by the results of the parametric analysis.

As well as analysing the comparisons of the five treatment classes, the non-parametric analysis provided information for the parametric analysis. This information was in the form of the determination of which instrument categories were important differentiators between successful and unsuccessful students and between above and below average (on achievement scores) students. This information was useful in the process of developing an analysis of co-variance and this is the topic of the next section: the parametric part of the statistical analysis.

5.3 The parametric analysis

5.3.1 An overview of the parametric analysis

Diagram 5.2 Flowchart of the parametric statistics used

<u>Matched control and treatment classes</u>

↓

Test for homogeneity of variance between control and treatment class

↓

Use: Bartlett's test on achievement scores to confirm a non-significant F value. Any non homogeneity is then due purely to chance and both control and treatment class can be assumed to come from the same population

↓

Calculate correlation between achievement scores and instrument categories. Check correlation for significance.

↓

If a significant correlation is found, proceed....

↓

Calculate homogeneity of regression

↓

If F is non-significant, proceed....

↓

Do an analysis of co-variance and construct regression equations that relate achievement scores to instrument categories

↓

Use matrices to determine the linear coefficients of co-variables

↓

Test the regression equations for accuracy

↓

The accuracy is checked by determining if the F value, in a comparison between predicted and actual achievement scores of selected treatment classes, is significant. A significant F indicates an accurate regression equation.

The objective of this second prong of attack is to formulate a regression equation relating achievement scores to several co-variables. One similarity between variables and co-variables is that both are variable quantities that can change over time. The main difference between them can be explained by looking at the characteristics of each. The change in a variable is caused by, or dependent on, a change in some other variable. Both the dependent and the independent variable are particular to the immediate environment in which the measurements take place. A co-variable, while changing with time, is not dependent on an independent variable which varies within the constraints of the immediate environment of the experiment. The following example illustrates the difference between variables and co-variables. In an educational experiment three different teaching methods are compared. The teaching method variables may be laboratory, expository and discovery. The teaching methods are the independent variables because they can be controlled. The dependent variable may be achievement scores. The students participating in the experiment will differ on some variables outside the experimenter's control. Although these variables can be measured on some scale, they are independent of the teaching method. These variables could be mathematical ability and reading ability. Because they cannot be controlled and yet have to be taken into account they can be treated as co-variables.

An analysis of co-variance relates the co-variable(s) to the dependent variable for each of the treatment and control variables. In educational parlance the dependent variable is called the criterion. The name of this dependent variable must not be confused with means by which the grade nine classes were assessed. These assessments were also called criteria. In this research the grade nine assessment criteria were represented as (end of quarter) achievement scores. The regression equations determined in this analysis had as the criterion the achievement score and as co-variables the instrument categories that were significant differentiators between successful and unsuccessful students. These instrument categories were co-variables in relation to achievement score because within the constraints of the experiment they were:

(1) unlikely to change in the time span of one quarter;
(2) unrelated to the method of teaching, at least in the short term;
(3) related to the students' level of cognitive development, the latter was related to time and other variables much more than to the method of teaching, especially in the short term;
(4) student characteristics which needed to be taken into account because they had an effect on achievement scores, as proven in the non-parametric analysis.

5.3.2 The analysis of co-variance and tests performed prior to the analysis of co-variance

This section will explain both the need for pre-tests and an analysis of co-variance and detail different statistical pre-tests. All calculations were performed on a hand held calculator because at the time the statistical software packages available did not have the particular calculations used in this study. As a result, the calculations are presented at a level of detail (F score calculations, setting up of matrices) appropriate to hand held calculators. The statistical tests used are as follows:

(1) test for homogeneity of variance;
(2) an analysis of variance;
(3) calculation of a correlation coefficient;
(4) test for homogeneity of regression;
(5) an analysis of co-variance;
(6) calculation of the regression equation.

Test for homogeneity of variance

Dayton (1970, 33) stated that for an ANOVA and for an analysis of co-variance to be valid it is important for the sample to be homogeneously distributed in the population's variance. A population that is normally distributed must have a homogeneous variance. However, a homogeneous variance does not guarantee a normally distributed population. Therefore, a test for population normality is also performed. If the variance of the sample is not normal, then the homogeneity of variance test reveals a significant value. That is, it is possibly highly skewed or bi-modal. If the test's value is non-significant than the difference in outcomes of the treatment and control groups is due solely to the different treatments the groups received. With regard to homogeneity and normality Dayton (1970, 35) stated that:

> There is a good deal of evidence that the analysis of variance is virtually unaffected by violations of normality and homogeneity of variance if the samples ...are of the same, or approximately the same, size.

The student population used in this research was tested both for normality and for homogeneity of variance using Bartlett's test (Dayton, 1970, 33).

An analysis of variance (ANOVA)

The assumptions contained in an analysis of variance, whether they are product-moment correlations, t-tests or, as in this ANOVA, F tests, are as follows (Nunnally, 1975, 281):

(1) data relate to interval or ratio scales;
(2) subjects are sampled from populations which are at least approximately normally distributed.

With regards to the shape of the distribution, there seems to be evidence to indicate that the validity of the ANOVA is not destroyed by a deviation from the normal distribution.

> All of these tests (p-m correlation, t-test, F-test) are "robust" in the sense that even marked violations of assumptions of normality of distributions have little influence on descriptive statistics or probabilities obtained from significance tests.
> (Nunnally, 1975, 281)

The reason for applying an ANOVA was to determine if there was a significant difference between treatment and control group on the basis of achievement scores. Dayton (1970, 313) states that there is relatively little point in applying an analysis of co-variance if the ANOVA does not show a significant difference. In other words, first find if there is a significant difference between treatment and control group in terms of achievement, then try to quantify what causes the differences.

The first ANOVA compared the average of the (achievement) criteria scores obtained by the students. Then, a series of ANOVA's was applied to different groupings of the separate criteria in order to find those criteria which caused the two groups to differ. Students in both groups were also compared on the different levels of cognitive development as determined by S.O.L.O. levels.

In ANOVA's F values are determined by comparing the mean squared deviations from mean between the two groups' mean as a ratio to the mean squared deviations from the mean of the two groups separately. That is, the Mean Squares between groups is divided by the Mean Squares within the groups. A significant F value indicates the difference between the groups is significantly greater than the way the scores differ between individual students within each group. A non-significant F value indicates the students vary more within each group than the scores differ between the two groups: there is no difference between the two groups.

Correlation coefficients between the instrument categories

It was judged unnecessary to calculate a correlation coefficient other than rho which had already been calculated. The reason for this is explained in the following argument.

From the rho values calculated, certain instrument categories were determined as non-significant differentiators between successful and unsuccessful students (see page 132). Even though these instrument categories may not have been significant in separating successful and unsuccessful students, it was possible that they may have been significant differentiators between the treatment and control group. However, the aim of this research was to determine the characteristics of students who coped well with the inquiry method of teaching science. Therefore, the choice of co-variable was made according to which characteristics differentiated between successful and unsuccessful students in the inquiry method and not which characteristics differentiated between the two treatment groups. After all, there was no point in finding out which characteristics were particular to inquiry teaching if those characteristics promoted unsuccessful students!

Homogeneity of regression

The method of calculating the homogeneity of regression was suggested by Dayton (1970, 323). Calculations of homogeneity of regression determine if the co-variables chosen exert an approximately equal effect on the criterion in terms of probability. This calculation is really a cross-check to confirm that correct co-variables have been selected. If, in fact, an unreasonable number of co-variables have no determining influence on the criterion, then this will show in a significant F value for the homogeneity of regression calculation. The criterion in the regression equation is the student achievement score.

Another cause for a significant F value is the possibility that the relationships between the criterion and the co-variables is not sufficiently linear to be able to calculate a regression equation. Thus, it is essential to have homogeneity of regression. This is indicated by a non-significant F value for the homogeneity of regression calculation (Dayton, 1970, 323). If a significant F value is obtained, the selection of co-variables has to be reviewed. This can be done by re-examining the correlation coefficients or by separately graphing each co-variable against the criterion to determine which co-variables are linearly related. Any non-linear relationships can then be discussed separately and a regression equation formulated with co-variables which are suited to a non-significant F value (Dayton, 1970, 323).

Although the calculation for homogeneity of regression is lengthy, the principle of construction of the F ratio is straight forward. F is the ratio of the average deviation in variance between the treatment and control group to the average deviation in variances of the co-variable scores of the two groups separately (Dayton, 1970, 323).

The criterion, Y, used in the homogeneity of regression calculations was not taken as the average of all assessment criteria scores of the students. Instead, a combination of selected criteria, on which the students were measured, was used. In the previous step in the parametric sequence, the ANOVA had indicated those criteria on which the most significant difference between treatment and control group existed. These were averaged and used as the criterion, Y.

A number of analyses of co-variance were performed. Table 5.3 details these analyses. For all analyses separate tests for homogeneity of regression had to be performed.

Table 5.3 Details of student samples and instrument categories used in the analyses of co-variance.

Student sample	Number of co-variables	Co-variables used in the analysis
treatment	2	Categorical-Class, Attitude
control	2	Categorical-Class, Attitude
combined	2	Categorical-Class, Attitude
treatment	6	Whole, Classify by Function, Categorical-Class, Attitude, Questioning-Recall, Principles-Application
control	6	Whole, Classify by Function, Categorical-Class, Attitude, Questioning-Recall, Principles-Application
treatment	9	Whole, Classify by Function, Categorical-Class, Attitude, Enjoyment, Questioning, Recall, Principles, Application

Analysis of co-variance

Only after all the foregoing steps have been satisfactorily performed can an analysis of co-variance be attempted. The first step in an analysis of co-variance determines if a significant relationship exists between the criterion (the achievement score) and the co-variables (see table 5.18). An F score is calculated in the same way as in an analysis of variance. If F is non-significant then any apparent relationship between criterion and co-variables is due purely to chance. This means that the co-variables can not be used to predict the criterion score. If the F score is significant then a relationship is not likely to be due to chance and the co-variables can be used to predict the criterion score. A regression equation can then be developed to show the relative influence of each of the co-variables on the criterion score.

In the previous paragraphs it was stated that a regression equation should only be developed if a significant F score were found. The actual calculations to derive an F score are extremely lengthy. An argument could be made in favour of developing a regression equation even if a non-significant F score was found. The argument relies on the thoroughness of all calculations performed prior to reaching the stage where an analysis of co-variance can be done. If all the intermediate calculations indicated a significant relationship exists between criterion and co-variable, then a non-significant F score can be assumed to be caused by some underlying factors that mask the significance of any relationship. Another reason for discounting a non-significant F score is based on the possibility that the effect on the F score of two opposite trends could be zero.

Therefore, even though two significant F scores were found, their significance was, in fact, not considered an imperative prerequisite for the construction of a regression equation. One F score was calculated for the data based on two co-variables. The other F score was calculated for the data based on six co-variables. These calculations discussed in the next section merely confirmed the significance calculations of the ANOVA performed previously.

Two methods for determining if the analysis of co-variance is significant

There are two possible ways to calculate an F score. One way is to use matrices (Dayton, 1970, 326-331) that help to calculate the sampling error in the criterion score. The other way (Dayton, 1970, 325) does not involve matrices and multiplies the mean square of the error in the criterion scores with a factor that includes the ratio of the sum of squares of the co-variable scores of the treatment group to the sum of squares of the errors in the co-variable scores. Both methods were used in this thesis. Both produced the same result.

With regard to the value of F in the analysis of co-variance, if the co-variables were chosen correctly, that is if they really were co-variables rather than variables, the F value should be more significant for the analysis of co-variance than for the ANOVA. This is due to the fact that the ANOVA ignores the co-variables and does not allow for variables that do not directly affect the criterion. In other words, the ANOVA is not sensitive enough. Dayton (1970, 325) was the only reference found that suggested this should be the case. Dayton did not state that if the F score in the analysis of co-variance is not as significant as in the ANOVA that the co-variables are wrongly chosen.

A method for calculating a regression equation

A number of possible regression equations can be built. Dayton (1970, 326) states that if a non-significant F score for homogeneity of regression was found, then the treatment and control groups can be assumed to have a linear relationship between achievement score (criteria) and instrument categories (co-variables). The coefficient of linearity (slope of the regression equation) describes the relationship between criterion and co-variables. Of course, even though they may have a linear relationship, they may be relatively displaced along the y-axis as determined by the means of the respective criterion scores.

In mathematical language, the common regression coefficient and the individual regression coefficients for treatment group and control group are calculated from the ratios of the sum of cross-products of deviation from the mean of the co-variables' cross-products and the sum of the squares of deviation from the individual co-variables (Dayton, 1970, 326).

This method can only be used with any degree of mathematical accuracy if it makes sense to combine the co-variable scores to calculate an average. This method can certainly be used if there is only one co-variable or when there are two co-variables and both co-variables measure the same characteristic. An example of the latter case would be the scale Categorical-Functional, in the Cognitive Style of Categorisation Behaviour instrument, and the scale Application, in the Combined Cognitive Preference Inventory instrument.

In all cases other than the cases mentioned in the previous paragraph it is best to use matrices to calculate separate coefficients of linearity for each of the co-variables, rather than combining the covariables into one overall coefficient. The use of matrices allows the calculations to keep track of more than one co-variable simultaneously rather than using averages of co-variables. Matrices also allow data of control and treatment groups to be pooled for an overall regression equation.

In short, the coefficients of linearity of each of the co-variables in the regression equation is based on the criterion scores, the co-variable scores, the cross-products among the co-variables and the cross-products of the co-variables and the criterion.

Explanation of the terms inverse and transpose matrix

This section is presented here specifically for calculations performed on hand held calculators. In the calculations two terms involving matrices are often used. These terms are: inversion and transpose. They must be very clearly set up as they mean two totally different things.

An inverse of a matrix is calculated by the method of row by row reduction. The inverse of a matrix is that matrix which, when multiplied by the original matrix, results in an identity matrix. The concept of the inverse matrix is necessary to allow a division to be converted into a multiplication (matrices cannot be divided). This method is explained by Thomas and Finney (1979, A-

5) and Washington (1985, 453). Another method of finding an inverse of a matrix is by the use of determinants as explained by Thomas and Finney (1979, A-13). Both methods were used in the calculations in this thesis. The determinant method is faster, especially in these calculations because the matrices are symmetrical about the main diagonal. This provides a good cross-check for the accuracy and correctness of the values calculated.

The term transpose is used to indicate that the rows of the original matrix must become the columns of the transpose matrix and, therefore, the columns of the original matrix become the rows of the transpose matrix. The transpose of a matrix is used to multiply incompatible matrices (inappropriately dimensioned) and to convert the product of two matrices into a single number.

Comments on the use of multiple co-variables

Matrices can be readily expanded to include any number of co-variables. It must be remembered, however, that with the inclusion of extra co-variables the length of the calculations increases as the multiple of a square. There are two causes for this. Because square matrices are included in the calculations the extra inclusion of one more co-variable results in the increase of the size of the square matrix by one dimension, that is one more column and one more row. This is due to the cross-products of co-variables among themselves and the cross-product of the new co-variable with the criterion.

With regard to the choice of the number of co-variables to be used in a regression equation account must be taken of the reduction in effect that each marginal increase of the number of co-variables has on the improvement in accuracy of the predicted criterion score. This is more generally known as the law of diminishing returns. Another factor to be taken into account is the loss of degrees of freedom in the error sums of squares used in the calculation of F. In this research the number of students involved in the study allowed the use of nine co-variables. Dayton (1970, 327) makes the following comment on the choice of the number of co-variables: 'in most research fields in education, it seems unlikely that more than three to five co-variables would be worthwhile in the sense of adding significantly to the multiple regression'.

<u>How to set up logical tables for the calculations of values needed in the matrices</u>

The outline of the calculations in the previous section, illustrated clearly that the data, both raw and calculated, must be set out logically. Muddled "house-keeping" could result in selecting and using the wrong data. In this research the data were set out as follows.

For each group, treatment and control, a table was set out with the following columns: each co-variable with the square of each co-variable next to each other, the criterion score with its square and a column for each of the cross-products (co-variable x co-variable and co-variable x criterion). The sum of each column was placed under the column. A third table was constructed next to summarize the sum totals of each column. This third table also contained a column showing sum totals of both groups added together. This way, whether calculating the Sum of Squares of Products and Sum of Squares values for each group separately or for all students as a whole, the data are immediately accessible.

5.4 The results of the statistical analyses

In the following sections all the results of the calculations are presented. The order of presentation is the same as that adopted in the previous section on the mathematical background of the analyses of the research data. Non-parametric statistical results are analysed first then the parametric analyses results are presented. Overall, the order is as follows.

The non-parametric analyses

(a) Mann-Whitney U test
(b) Chi-square
(c) Guttman's lambda
(d) Kendall's W
(e) Inter-instrument categories' correlation
(f) W.N.P.D.

The parametric analyses

(a) Test for homogeneity of variance
(b) ANOVA
(c) Correlation coefficients between instrument categories
(d) Test for homogeneity of regression
(e) Analysis of co-variance, including significance
(f) Regression equations

5.4.1 The non-parametric analyses

<u>(a) Mann-Whitney U test</u>

The results of comparing unsuccessful and successful students on the basis of achievement scores are quantified by the U scores. The descriptors successful and unsuccessful were chosen to indicate whether students coped well or poorly with the inquiry method. Definitions of successful and unsuccessful were given in chapter 3 section 3.5.2 (see page 67). Students were first classified as unsuccessful or successful based on teacher observations and a teacher's daily logbook. This classification was later validated by the results of the S.P.C.S. instrument (see page 75). The U scores are converted to z scores and can be analyzed for significance in tables. The comparison is by rank. Therefore, if the average rank of unsuccessful students is higher than the average rank of successful students, then the successful students performed better!

Table 5.4 The results of the Mann-Whitney U test on the five inquiry classes.

Inquiry class	average rank	significance
Bubbles and Bangs	Uav> Sav	p< 0.016
Essential Chemistry	Uav> Sav	only U significant
Green Machine	Uav> Sav	p< 0.001
Chemistry in the Market	Uav> Sav	p< 0.048
Advanced Science	Uav> Sav	only U significant

Legend: U= unsuccessful students
S= successful students
av= average
p= probability of student scores not being significant
only U significant= only the distribution of ranks of unsuccessful students
was significantly different from the predicted frequencies.

Table 5.4 shows that in all treatment classes successful students outperformed unsuccessful students. In three classes successful students performed significantly better than predicted and unsuccessful students performed significantly worse. In the Essential Chemistry and Advanced Science classes unsuccessful students performed significantly worse than predicted, but successful students did not perform significantly different than predicted.

As part of this step the distribution of students on S.O.L.O. categories was determined. From these data, the students could be compared to the data obtained by Collis and Davey (1984, 15).

Table 5.5 Distribution of students on S.O.L.O. levels in all participating classes.

Class	Grade	S.O.L.O. Category							
		Singular		Multiple		Relational		General Abstract	
		number	%	number	%	number	%	number	%
Bubbles and Bangs	9	0	0	19	86	3	14	0	0
Essential Chemistry	9	0	0	9	56	6	38	1	6
Grade 9 Total	9	0	0	28	74	9	24	1	2
Green Machine	10	0	0	9	50	9	50	0	0
Chemistry in the Market	10	0	0	9	64	5	36	0	0
Advanced Science	10	0	0	0	0	11	73	4	27
Grade 10 Total	10	0	0	18	38	25	53	4	9
Total (85 students)		0	0	46	54	34	40	5	6

Collis and Davey's data (1984, 18) on the actual S.O.L.O. instrument items used in this research average to S.O.L.O. Multiple 60.4% and S.O.L.O. Relational 25.4% for grade nine students. This compares to a grade nine average of 74% at S.O.L.O. Multiple and 24% at S.O.L.O. Relational found in this research. It appears the data used in this research are similar to the data used by Collis and Davey.

(b) Chi-square test results

This test investigated the significance of the difference between successful and unsuccessful students in two instruments: the Combined Cognitive Preference Inventory (C.C.P.I.) and the Cognitive Styles of Categorization Behaviour instrument (C.S.C.B.). The students were compared on the basis of their scores being above the mean, one standard deviation above the mean or one standard deviation below the mean. These scores relate to all categories within the two instruments. If the Chi-square value was significant then there was a relationship between being either successful or unsuccessful and the instrument scores.

Table 5.6 Level of significance and the basis of comparison of the difference between successful and unsuccessful students for the Combined Cognitive Preference Inventory and Cognitive Styles of Classification Behaviour instruments.

Instrument	basis of comparison	level of significance
C.C.P.I.	one standard deviation above mean	non-significant
C.C.P.I.	one standard deviation below mean	non-significant
C.C.P.I.	above the mean	$p < 0.016$
C.S.C.B.	one standard deviation above mean	$p < 0.05$
C.S.C.B.	above the mean	$p < 0.02$

It appears from table 5.6 that successful and unsuccessful students could be statistically separated on the basis of above average scores. This meant that, in terms of coping with the inquiry method, successful and unsuccessful students could be separated on bases other than ability and end of quarter achievement scores. The bases on which these students could be separated were on above average scores on all categories in both instruments (C.C.P.I. and C.S.C.B.).

(c) Guttman's lambda

In this test the strengths of the relationships between successful/unsuccessful and the different instrument categories were given numerical values called lambda. The greater the value of lambda the stronger the relationship. The Chi-square significance test results were used. Above average instrument category scores were used for both C.C.P.I. and C.S.C.B. and, also, one standard deviation above the mean scores were used for C.S.C.B.

Table 5.7 Strength of the relationship between the successful and unsuccessful student classifications based on above average C.C.P.I. instrument category scores.

S.O.L.O. category	C.C.P.I. category	lambda
Multiple	all categories	0.028
Relational	all categories	0.032
Multiple	Principles, Application, Questioning-Recall, Principles-Application	0.118
Multiple + Relational	all categories	0.031
Multiple + Relational	all categories, except Recall	0.033
Multiple + Relational	Principles, Application, Questioning-Recall, Principles-Application	0.143
all students	all categories	0.040
all students	Principles, Application, Questioning-Recall, Principles-Application	0.029

From table 5.7, it appears that the C.C.P.I. categories Principles, Application, Questioning-Recall and Principles-Application provide the best means of separating successful and unsuccessful students. This result appears to apply only to S.O.L.O. Multiple and Relational students and not to S.O.L.O. Singular and General Abstract students. These latter S.O.L.O. students only comprise 0% and 6% of all students respectively.

Table 5.8 Strength of the relationship between the successful and unsuccessful student classifications based on above average C.S.C.B. instrument category scores.

S.O.L.O. category	C.S.C.B. category	lambda
all students	all categories	0.057
all students	Part-whole, Whole, Categorical-Functional Categorical-Class	0.052
all students	Whole,Categorical-Class, Categorical-Functional	0.066
all students	Part-whole, Categorical-Class, Categorical-Inferential	0.081
Multiple	all categories	0.030
Relational	all categories	0.046

Table 5.8 shows that out of all C.S.C.B. categories the categories Part-whole, Categorical-Class and Categorical-Inferential separated successful and unsuccessful students best. The categories Whole and Categorical-Functional were the next best separators.

The strength of the separation between successful and unsuccessful students based on C.S.C.B. categories was not as strong as that for some C.C.P.I. categories.

Table 5.9 Strength of the relationship between the successful and unsuccessful student classifications based on one standard deviation above the mean C.S.C.B. instrument category scores.

S.O.L.O. category	C.S.C.B. category	lambda
all students	all categories	0.132
all students	all categories, except Functional-Interdependence	0.143
all students	Part-whole, Whole, Categorical- Functional, Categorical-Class	0.162
all students	Whole, Categorical-Functional, Categorical-Class	0.170

It appears form table 5.9 that the differences between successful and unsuccessful students were much stronger for students scoring one standard deviation above the mean than for students scoring above the mean. The categories Whole, Categorical-Functional and Categorical-Class were the strongest separators of successful and unsuccessful students.

To provide more detail about lambda and in order to show more clearly which instrument categories were instrumental in measuring difference between successful and unsuccessful students, the following tables were constructed. There are three tables. Each table corresponds consecutively to one of the three tables above. It should be remembered that if Unsuccessful>Successful then the average unsuccessful rank is numerically greater than the average successful rank. This would mean that unsuccessful students did not perform as well as successful students. For clarity, this situation will be described as successful students outscoring unsuccessful students.

Table 5.10 The values of lambda and the direction of the relationship (on a rank basis) between successful and unsuccessful students for six categories of the C.C.P.I. instrument.

	C.C.P.I. instrument (above the mean scores)			
Instrument category	lambda=0.143 U<or>S	lambda=0.118 U<or>S	lambda=0.040 U<or>S	lambda=0.032 U<or>S
Recall			>	=
Principles	>	>	>	=
Questioning			<	<
Application	<	<<	<<	=
Questioning-Recall	<	>	<	<<
Principles-Application	>	>	>	<

Legend: <<=much smaller than,
U<or>S=average ranking of unsuccessful students is either greater, >, or smaller, <, than average ranking of successful students.

Note: these data have been weighted, that is a correction has been made for the fact that there were fewer Unsuccessful students than Successful students.

The categories of the C.C.P.I. instrument were chosen according to the strengths of the relationships. These strengths were expressed as lambda. Six categories were chosen: Recall, Principles, Questioning, Application, Questioning-Recall and Principles-Application. Table 5.10 shows that, with the exception of Questioning-Recall, the direction of the relationships was consistent for most of the values of lambda. Thus, in the category Recall successful students outscored (ranked better), or equalled, unsuccessful students. In category Principles, successful students outscored unsuccessful students. In the categories Questioning, Application and Questioning-Recall unsuccessful students scored higher than successful students. In the category

Principles-Application successful students scored higher than unsuccessful students.

Table 5.11 The values of lambda and the direction of the relationship (on a rank basis) between successful and unsuccessful students for six categories of the C.S.C.B. instrument.

Instrument category	C.S.C.B. instrument (above the mean scores) lambda=0.081 U<or>S	lambda=0.066 U<or>S	lambda=0.057 U<or>S	lambda=0.052 U<or>S
Part-whole	>		>	>
Whole		>	>	>
Functional-Interdependence			>	
Categorical-Functional		<	<	<
Categorical-Class	<	<	<	<
Categorical-Inferential	<<		<<	

Legend: <<=much smaller than,
U<or>S=average ranking of unsuccessful students is either greater, >, or smaller, <, than average ranking of successful students.

Note: these data have been weighted, that is a correction has been made for the fact that there were fewer Unsuccessful students than Successful students.

The strengths of the relationships for the C.S.C.B. instrument were not as strong as for the C.C.P.I. instrument as judged by the values of lambda. The direction of the relationships is perfectly consistent at all values of lambda. Successful students outscored unsuccessful students on three categories: Part-whole, Whole and Functional-Interdependence. Successful students scored lower than unsuccessful students on the three other categories: Categorical-Functional, Categorical-Class and Categorical-Inferential.

Table 5.12 The values of lambda and the direction of the relationship (on a rank basis) between successful and unsuccessful students for six categories of the C.S.C.B. instrument based on scores one standard deviation above the mean.

	C.S.C.B. instrument (one standard deviation above the mean)			
Instrument category	lambda=0.170 U<or>S	lambda=0.162 U<or>S	lambda=0.143 U<or>S	lambda=0.132 U<or>S
Part-whole		>	>	>
Whole	>	>	>	>
Functional-Interdependence				=
Categorical-Functional	<	<	<	<
Categorical-Class	<<	<<	<<	<<
Categorical-Inferential			<	<

Legend: <<=much smaller than,
U<or>S=average ranking of unsuccessful students is either greater, >, or smaller, <, than average ranking of successful students.

The strengths of the relationships between successful and unsuccessful students were much stronger when based on one standard deviation above the mean scores than above the mean scores. The direction of the relationships were identical in both tables 5.11 and 5.12.

<u>(d) Kendall's W</u>

In this calculation the value of W reveals if the relationship between the criterion (achievement) score and the instrument category under investigation is significant, to a level of $p<0.05$. The calculations were performed on a class by class basis. First, all the instrument categories were included in the calculation of W. Then, by a process of elimination those categories that were judged to contribute least to the significance of W were excluded and W was re-calculated until a significant W was found. This process of elimination was based on Spearman's rank order correlation coefficient, rho, between the instrument categories and end of quarter achievement scores. Categories with the lowest rho values were eliminated first. Only the significant W relationships are presented.

Table 5.13 Instrument categories which were significantly related to criterion (end of quarter achievement) scores as determined by Kendall's W.

Class (Grade)	Significant instrument categories
Bubbles and Bangs (9)	Categorical-Functional, Attitude, Questioning- Recall, Principles- Application.
Experimental (9)	Whole, Categorical-Functional, Categorical-Class, Attitude, Enjoyment, Principles, Application, Questioning-Recall, Principles-Application.
Green Machine (10)	Categorical-Functional, Attitude, Questioning-Recall, Principles-Application.
Chemistry in the Market (10)	Attitude, Principles-Application.
Advanced Science (10)	Principles-Application, Questioning-Recall, Attitude

Note: the description significant refers to the fact that the value of W was significant at a confidence level of $p<0.05$.

Comparing tables 5.12 and 5.13 based on Guttman's lambda with those based on Kendall's W respectively, several observations can be made. From the six C.C.P.I. categories only four remain as significant differentiators between successful and unsuccessful students: Principles, Application, Questioning-Recall and Principles-Application. From the six C.S.C.B. categories only three remain: Whole, Categorical-Functional and Categorical-Class. The T.O.S.R.A. instrument contributed only two categories which significantly differentiated between successful and unsuccessful students: Attitude and Enjoyment. In the next section these seven significant categories are correlated to achievement score.

(e) Inter-instrument category and end of quarter achievement score correlation data

The correlation calculations performed produced rho values between different instrument categories as well as between instrument categories and achievement scores. The correlation data can be considered important and significant because all the instrument categories used had significant relationships with both unsuccessful and successful student categories (Guttman's lambda and Chi-square) and achievement score (Kendall's W).

Table 5.14 Correlations between selected instrument categories for each of the treatment classes.

Bubbles and bangs (grade 9)

W	CF	CC	At	E	P	A	Q-R	P-A	
-.36	+.44	+.20	+.39	+.26	-.14	+.39	-.23	-.35	%
	-.63	-.29	+.10	-.38	+.18	-.46	-.01	+.42	W
		+.33	+.07	+.23	-.39	.30	.13	-.47	CF
			.12	-.16	-.26	.52	-.18	-.50	CC
				.43	.09	.03	-.24	-.04	At
							.29	-.19	E
						.33	-.49	.78	P
							-.11	-.83	A
								-.19	Q-R

Essential Chemistry (grade 9)

W	CF	CC	At	E	P	A	Q-R	P-A	
.02	.03	.38	.34	.44	.05	.03	-.48	-.01	%
	-.39	-.27	-.06	-.54	-.03	.04	-.06	-.15	W
		.05	.40	.25	-.20	.37	-.18	-.59	CF
			.35	.41	.25	-.08	-.33	.22	CC
				.75	.17	.07	-.67	-.02	At
					.17	-.10	-.54	.23	E
						.41	-.59	.14	P
							-.25	-.79	A
								.02	Q-R

Green Machine (grade 10)

W	CF	CC	At	Q-R	P-A	
-.14	.15	.16	-.27	-.16	.19	%
			-.21	-.17	.30	CF
				-.02	.23	At
					-.42	Q-R

Chemistry in the Market (grade 10)

W	CF	CC	At	Q-R	P-A	
.00	-.20	-.13	-.41	.14	.55	%
			.25	-.11	.16	CF
				.06	-.35	At
					.04	Q-R

Advanced Science (grade 10)

CF	At	Q-R	P-A	
-.18	-.51	.59	.25	%
	-.18	-.07	-.59	CF
		-.77	-.08	At
			.47	Q-R

Legend: % = end of quarter achievement score
W =Whole
CF =Categorical-Functional
CC =Categorical-Class
At =Attitude to science
E =Enjoyment of science
P =Principles
A =Application
Q-R =Questioning-Recall
P-A =Principles-Application

The correlations between the different instrument categories were calculated on a class by class basis in order to keep the variables of science topic and student ability constant. It appears (table 5.14) that not only did the absolute value of the correlation differ from class to class, the sign of the correlation also varies from class to class. No general patterns were apparent in the data.

(f) The Weighted Net Percentage Difference calculations

The instrument categories that were included in the W.N.P.D. table were those categories that were shown to have significant relationships with both the successful/unsuccessful classification and with the criterion score. These relationships were established in the foregoing calculations (Guttman's lambda, Chi-square and Kendall's W). The W.N.P.D. calculations did not use actual student numbers but used percentage of students who fitted in the cells of the W.N.P.D. table. For instance, of a maximum of 18 students only eight, or 44%, could be classified in the first cell of the second row. Table 5.15 below shows percentages of students, relative to the maximum possible, who fitted in the cells are shown.

Table 5.15 will be discussed in detail in the next section (see page 165). In general, it can be seen that none of the students who participated in the experiment rated S.O.L.O. Singular. Also, none of the unsuccessful students were rated S.O.L.O. General Abstract.

Examples of comparisons that were made using the W.N.P.D. table

A total of 145 comparisons were made. An example of one comparison was: how many more times was a successful student more likely to score above the criterion score average than below it? This compared a block of the first 36 cells (above average successful students) to another block of 36 cells: cells 73 through to 108 inclusive (below average successful students). The actual comparison only compared equivalent cells with each other. Thus, cell 1 was compared with cell 73. Cell 1 was the above average successful student (rated Singular on the S.O.L.O. instrument) and scored above average on the Questioning-Recall category. Cell 73 was the below average Successful student (rated Singular on the S.O.L.O. instrument) and scored above average on the Questioning-Recall category. Cells 1 to 9 were all zero, as were cells 73 to 81; therefore, these cells did not need to be considered. The actual W.N.P.D.

calculation was explained by Cohen (1976, 379). The result of the calculation showed a W.N.P.D.= -1.71%; that is, Successful students were 1.7% more likely to score below the average than above it based on all the instrument categories.

Table 5.15 Percentage of students in the W.N.P.D. cells.

Criterion score	classified as	SOLO level	Q-R x>x	Q-R x<x	P-A x>x	P-A x<x	W x>x	CF x>x	CC x>x	At x>x	E x>x
		S	0	0	0	0	0	0	0	0	0
		M	44	50	50	50	47	65	65	61	72
	successful	R	47	53	71	29	35	41	41	59	65
		G	67	33	67	33	33	0	100	67	33
above the criterion mean											
		S	0	0	0	0	0	0	0	0	0
		M	25	75	75	25	20	80	60	40	0
	unsuccessful	R	43	43	43	43	50	50	67	17	33
		G	0	0	0	0	0	0	0	0	0
		S	0	0	0	0	0	0	0	0	0
		M	33	67	44	56	45	45	55	80	50
	successful	R	75	25	75	25	75	25	50	100	50
		G	0	100	0	100	50	50	100	0	100
below the criterion mean											
		S	0	0	0	0	0	0	0	0	0
		M	67	33	78	22	60	30	50	40	40
	unsuccessful	R	25	100	75	50	50	33	33	60	60
		G	0	0	0	0	0	0	0	0	0

Legend: x<x = student's score < average score of all students in that category
SOLO = SOLO classification
criterion score= either higher than the average or below the average criterion score. Average of all classes as a whole.

Note: unsuccessful and successful classification originally based on teacher observations and later confirmed by significant Mann-Whitney U values and S.P.C.S. instrument results.

A similar calculation for unsuccessful students resulted in a W.N.P.D. value of -6.49%. The calculations were modified to include only specific instrument categories. If, for instance, the comparison had been formulated as: how many more times was a successful student more likely to score above the criterion score average than below it, based on above average scores in all instrument categories (exclude Questioning-Recall below the category mean scores and Principles-Application below the category mean scores), the W.N.P.D. value would have been 0.13%.

As well as a block by block comparison, a row by row comparison could also be made by stipulating a specific S.O.L.O. level. In total, well over a hundred comparisons were made to determine to what extent successful students were different from unsuccessful students and on which instrument categories. Also, comparisons were made to find out the extent to which, for instance, Questioning-Recall above the category mean score students were different from Questioning-Recall below the category mean score students.

Significance testing of the trends in the W.N.P.D. table

The percentages themselves cannot be examined for significance. One way of overcoming this problem is to group the W.N.P.D. comparison results in some way and then examine if this grouping exhibits a significant trend. One example of this method is a test for significance using the Mann-Whitney U test outlined in the section on non-parametric statistics (see page 109). A trend that was investigated, among many others, asked if there was a significant difference between successful and unsuccessful students who scored above average on criterion scores on the basis of above average Questioning-Recall score. There was a significant difference, $p<0.04$, with successful students performing significantly better. The same was found for scientific Attitude, $p<0.04$, and for Enjoyment of science, $p<0.07$, but not for any other instrument categories. Similar significance tests were done on above average successful to below average successful students, idem for unsuccessful students, comparisons on the basis of S.O.L.O. classifications. Example calculations are given in Appendix B. One of the results of the Guttman's lambda calculations found that unsuccessful inquiry students are more likely to score above average in the category Questioning-Recall than successful inquiry students. This result is not contradicted by the statement that: eventhough above average achievement students are less likely (9%) to score above average in the Questioning-Recall category (compared with below average achievement students), of the above average achievement students that go against this trend the successful inquiry students outperform the unsuccessful inquiry students in achievement.

Summary of significance calculations

Many calculations, like the example in the previous section, were performed. A summary of these calculations is presented in this section. This summary should be read in conjunction with the W.N.P.D. data, because the significance values lend strength to the W.N.P.D. values.

In tables 5.16 and 5.17 below the descriptions: 'better' and 'lower', in the second column of each table, refer to the average ranking of the designated group. For example, "Q-R above mean S above mean better than U above mean" means that the group of students who scored above the mean in the category Questioning-Recall is being considered. In this group the successful (above the criterion average scoring) students' average rank was significantly better than the average rank of unsuccessful (above the criterion average scoring) students. That is, here successful students outperformed unsuccessful students.

Table 5.16 Significant relationships between successful and unsuccessful students for different instrument categories.

Instrument category	Relationship and S.O.L.O. level
Q-R above mean	S above mean better than U above mean, all students
P-A above mean	S above mean lower than U below mean, all students
P-A above mean	U above mean better than (U+S) above mean, S.O.L.O. Multiple and Relational
P-A above mean	U below mean better than S above mean, S.O.L.O. Multiple and Relational
W above mean	S above mean lower than U below mean, all students
W above mean	(U+S) above mean lower than (U+S) below mean, all students
W above mean	U below mean better than S above mean, S.O.L.O. Multiple and Relational
CF above mean	S above mean better than U below mean, all students
CF above mean	(U+S) above mean better than (U+S) below mean, all students
CF above mean	U above mean lower than (U+S) above mean, S.O.L.O. Multiple and Relational
CF above mean	U below mean lower than S above mean, S.O.L.O. Multiple and Relational
CC above mean	(U+S) above mean better than (U+S) below mean, all students
CC above mean	U above mean lower than (U+S) above mean, S.O.L.O. Multiple and Relational
At above mean	S above mean better than U above mean, all students
At above mean	(U+S) above mean lower than (U+S) below mean, all students
E above mean	S above mean better than U above mean, all students
E above mean	S above mean better than U below mean, all students
E above mean	S above mean better than (U+S) below mean, all students
E above mean	U below mean lower than S above mean, S.O.L.O. Multiple and Relational

Legend: Q-R=Questioning-Recall (C.C.P.I.)
P-A=Principles-Application (C.C.P.I.)
W=Whole (C.S.C.B.)
CF=Categorical-Functional (C.C.P.I.)
CC=Categorical-Class (C.C.P.I.)
At=Attitude to science (T.O.S.R.A.)
E=Enjoyment of science (T.O.S.R.A.)

U=unsuccessful inquiry students
S=successful inquiry students

In table 5.17 below a dual comparison is undertaken. By this is meant a W.N.P.D. comparison is contrasted against another W.N.P.D. comparison. The question that was asked in the formulation of the first of the thirteen comparisons was: is there a significant relationship between above average successful and unsuccessful students on the one hand and students in general scoring above and below the average criterion score, on the basis of the categories Questioning-Recall above the mean, Principles-Application above the mean, Whole, Categorical-Functional, Categorical-Class, Attitude, Enjoyment? The answer to this question was: yes, a significant ($p<0.04$) correlation [-0.56] exists. This meant that, essentially, comparing above average successful and unsuccessful students was the same as comparing above to below average students with the important distinction that the W.N.P.D. values had the reverse signs attached to them. For example, successful students scoring above mean in Questioning-Recall are 10% more likely to score above the mean than unsuccessful students and, in general, students scoring above average in Questioning-Recall are 10% less likely to score above the average than below the average criterion score (note the negative correlation). As explained previously (refer page 166), most students who score above average in Questioning-Recall tend to be unsuccessful inquiry students. Therefore, this example again indicates that successful inquiry students, who go against the trend and score above average in the Questioning-Recall category, outperform unsuccessful inquiry students.

Table 5.17 shows very high, significant relationships between the various dual comparisons. The last dual comparison is a virtual identity.

Table 5.17 Correlation results of the dual comparisons.

Categories being compared	categories, [rho value], (significance)
1(S vs U above mean) versus (above mean vs below mean)	Q-R above mean, P-A above mean, W, CF, CC, At, E, [-0.56], (p<0.04)
2(S vs U above mean) versus (above mean vs below mean)	Q-R above mean, P-A above mean W, CF, CC, At, [-0.90], (p<0.01)
3(S above mean vs S below mean) versus (above mean vs below mean)	Q-R above mean, P-A above mean W, CF, CC, At,E, [0.81], (p<0.01)
4(S above mean vs S below mean) versus (above mean vs below mean)	Q-R above mean, P-A above mean W, CF, CC, At, [0.93], (p<0.01)
5(S above mean vs S below mean) versus (above mean vs below av)	Q-R above mean, W, CF, CC, At, E, [0.95], (p<0.01)
6(U above mean vs U below mean) versus (above mean vs below mean)	Q-R above mean, P-A above mean, W, CF, CC, At,E, [0.76], (p<0.01)
7(U above mean vs U below mean) versus (above mean vs below mean)	Q-R above mean, P-A above mean, W, CF, CC, At, [0.89], (p<0.01)
8(S above mean vs U below mean) versus (above mean vs below mean)	Q-R above mean, P-A above mean, W, CF, CC, At,E, [0.43], (p<0.08)
9(S above mean vs U below mean) versus (above mean vs below mean)	Q-R above mean, P-A above mean, W, CF, CC, At, [0.45], (p<0.08)
10(S above mean vs U below mean) versus (above mean vs below mean)	Q-R above mean, P-A above mean, W, CF, CC, [0.89], (p<0.01)
11(U above mean vs S below mean) versus (above mean vs below mean)	Q-R above mean, P-A above mean, W, CF, CC, At,E, [0.87], (p<0.01)
12(U above mean vs S below mean) versus (above mean vs below mean)	Q-R above mean, P-A above mean, W, CF, CC, At, [0.95], (p<0.01)
13(U above mean vs S below mean) versus (above mean vs below mean)	Q-R above mean, W, CF, CC, At, [0.99], (p<0.01)

Legend: S=Successful students
U=Unsuccessful students
vs=versus
Q-R=Questioning-Recall (C.C.P.I.)
P-A=Principles-Application (C.C.P.I.)
W=Whole (C.S.C.B.)
CF=Categorical-Functional (C.S.C.B.)
CC=Categorical-Class (C.S.C.B.)
At=Attitude to science (T.O.S.R.A.)
E=Enjoyment of science (T.O.S.R.A.)

5.4.2 The parametric analysis

Before an investigation into the co-variables could begin, the criterion, to which these co-variables will be related, had to be precisely defined. Because the grade nine criteria consist of ten different criteria a decision had to be made regarding the selection of criteria to be used in the analysis of co-variance. At first, the average of all ten criteria was considered. Then, many calculations were performed to find out on which criteria the experimental group differed most from the control group.

<u>Bartlett's test for homogeneity of variance on criterion average scores</u>

The calculations performed followed the technique outlined by Dayton (1970, 33). For both the experimental and the control group the mean criterion score, the squared deviation from the average and, hence, the error sum of squares were calculated. The resulting Chi-square value was non-significant and the ANOVA could continue.

The ANOVA's F value between experimental and control groups was 0.63. This value was non-significant. This meant that the groups varied more within group than they varied between groups. In other words, the experimental and control groups did not differ significantly on the basis of average criterion scores. To achieve a significant F value the criterion score used in the analysis of co-variance had to be a limited selection of the ten criterion scores instead of an average of all ten.

In order to check that the control and experimental groups did perform differently, the control group used above was compared to a different, matching class taught during quarter two. This class was not considered part of the experiment in the first instant but was taught exactly the same content as the control and experimental groups. A three way comparison between quarter two, control and experimental groups on the groups' averages on all ten criterion scores separately revealed that the experimental group outperformed the matched quarter two and control groups in criteria Prl, Cl, I3 and K. It will be seen later that these criteria are the same as determined by significance calculations comparing just the experimental and control groups.

A series of F calculations was performed to determine which combination of the ten criteria showed the difference between experimental and control groups most. Table 5.18 below shows criteria combinations, the F value that represented the statistical significance of the difference between control and experimental groups, the level of probability of the F value and, finally, whether the average score of the experimental group was greater or smaller than the average score of the control group.

The four criteria on which the experimental group always outscored the control group were Pr-1, C-1, I-3 and K. For top level students (S.O.L.O. levels General Abstract and transitional Relational-General Abstract) the experimental group always outscored the control group. In S.O.L.O. level Relational the experimental group outscored the control group on criteria Pr-1, C-1, I-3 and K; but there was no difference on criteria Pr-2, Pr-3, I-3 and I-1. In S.O.L.O. level Multiple the experimental group outscored the control group on criteria Pr-1, C-1, I-3 and K; but was outscored by the control group on criteria Pr-2, Pr-3 and I-4.

Because the experimental group outscored the control group on criteria Pr-1, C-1, I-3 and K at any S.O.L.O. level, the average of these four criteria was used as the achievement score in further analyses.

Table 5.18 The effects on F value, statistical significance (p) and relative performance of experimental and control groups caused by different combinations of end of quarter criterion scores.

SOLO	Criterion combination										F	p<	Te?Tc
	Pr1	Pr2	Pr3	C1	C2	I1	I2	I3	I4	K			
all students	Pr1	Pr2	Pr3	C1	C2	I1	I2	I3	I4	K	0.63	0.093	>
all students	Pr1			C1		I1					0.05		
all students	Pr1			C1				I3		K	0.52	0.095	>
all students		Pr2	Pr3			I1			I4		0.86	0.09	<
all students		Pr2	Pr3					I3	I4	K	0.38	0.096	<
all students		Pr2	Pr3						I4		1.03	0.090	<
all students		Pr2	Pr3		C2				I4		0.58	0.094	<
all students		Pr2	Pr3								1.70	0.080	<
all students	Pr1			C1						K	0.33		>
all students				C1						K	0.32		>
all students				C1							0.24		
all students	Pr1	Pr2	Pr3	C1				I3		K	0.00		
Multiple		Pr2	Pr3								0.30	0.094	<
Multiple	Pr1			C1				I3		K	0.56	0.092	>
Relational +R/G		Pr2	Pr3								0.16		
Relational +R/G	Pr1			C1				I3		K	3.23	0.06	>
Relational		Pr2	Pr3								0.00		
R/G +General Abstract		Pr2	Pr3								1.78	0.078	>
R/G +General Abstract	Pr1			C1				I3		K	1.77	0.078	>

Legend: R/G=transition Relational to General Abstract
Pr1=Practical 1 (Tasmanian Certificate of Education criterion)
Pr2=Practical 2 (Tasmanian Certificate of Education criterion)
Pr3=Practical 3 (Tasmanian Certificate of Education criterion)
C1=Communication 1 (Tasmanian Certificate of Education criterion)
C2=Communication 2 (Tasmanian Certificate of Education criterion)
I1=Investigating 1 (Tasmanian Certificate of Education criterion)
I2=Investigating 2 (Tasmanian Certificate of Education criterion)
I3=Investigating 3 (Tasmanian Certificate of Education criterion)
I4=Investigating 4 (Tasmanian Certificate of Education criterion)
K=Knowledge (Tasmanian Certificate of Education criterion)
Te?Tc=Experimental criterion score average >, < or equal to Control Group
> or <=better than or lower than, respectively

The selection of co-variates

Having selected the appropriate achievement criteria used as the criterion in the analysis of co-variance, the next step was to select the appropriate co-variates. The first place to start the process was to look at the correlation data of experimental and control groups relating achievement scores to instrument categories. From these instrument categories a selection was made to obtain the most representative regression equation.

Table 5.19 Correlation data among achievement scores and instrument categories.

W	CF	CC	At	E	Q-R	P-A	
0.02	0.03	0.38	0.34	0.44	-0.48	-0.01	% experimental
0.04	-0.18	0.00	-0.11	0.67	-0.26	0.69	% control
	-0.39	-0.27	-0.06	-0.54	-0.06	-0.15	W experimental
	-0.45	-0.11	0.06	0.23	0.25	0.32	W control
		0.05	0.40	0.25	-0.18	-0.59	CF experimental
		0.03	0.11	0.04	-0.41	0.02	CF control
			0.35	0.41	-0.33	0.22	CC experimental
			0.05	-0.03	0.05	-0.43	CC control
				0.75	-0.67	-0.02	At experimental
				0.40	-0.09	-0.04	At control
						0.23	E experimental
						0.40	E control

Legend: W=Whole
CF=Categorical-Functional
CC=Categorical-Class
At=Attitude to science
E=Enjoyment of science
Q-R=Questioning-Recall
P-A=Principles-Application
%= achievement score

From these data the most promising instrument categories were selected based on correlation data. The following instrument categories were selected as co-variates: Categorical-Class, Attitude, Enjoyment, Questioning-Recall, Principles-Application. The question that had to be answered now was: are these correlation data significant? The calculations used to answer this question were based on the calculations suggested by Cohen (1976, 332):

sigma $r_0=(N-1)^{-.5}$; t-score= rho/sigma r_0 ; df= N-1

Table 5.20 Significance calculation for correlation data of the experimental group.

% versus	CC	At	E	Q-R	P-A
sigma r_0	0.2236	0.2085	0.2085	0.2182	0.2182
t-score	1.6995	1.6307	2.1103	-2.1998	-0.0458
df	20	23	23	21	21
p<	0.0526	0.0602	0.0236	-0.0206	-0.2588
significant	yes	yes	yes	yes	no

Table 5.21 Significance calculation for correlation data of the control group.

% versus	CC	At	E	Q-R	P-A
sigma r_0	0.2236	0.2236	0.2236	0.2182	0.2182
t-score	0	-0.4919	2.9964	-1.1916	3.1622
df	20	20	20	21	21
p<	1.0000	-0.2035	0.0041	-0.1161	0.0030
significant	no	no	yes	no	yes

Legend: df=degrees of freedom
p<=level of significance
CC=Categorical-Class
At=Attitude to science
E=Enjoyment of science
Q-R=Questioning-Recall
P-A=Principles-Application
%= achievement score

All five instrument categories seemed potential co-variates because all five were significant in at least one group (see tables 5.20 and 5.21). As a trial run the first analysis of co-variance was performed with the co-variates Categorical-Class and Attitude. In all, six regression equations were calculated. All regression equations used the average of achievement criteria Pr-1, C-1, I-3 and K as the criterion. The number and choice of instrument categories as co-variates varied to achieve more accurate regression equations.

Table 5.22 Co-variates used in the six regression equations calculated.

Number of co-variables	Co-variables used	Regression equation calculated applicable to:
2	Categorical-Class, Attitude	all students, experimental group, control group
6	Whole, Categorical-Functional, Categorical-Class,Attitude, Questioning-Recall, Principles-Application	all students, experimental group
9	Whole, Categorical-Functional, Categorical-Class, Attitude, Enjoyment, Questioning, Recall, Principles, Application	all students

Three regression equations with co-variates CC and At

These three equations served as a trial run to observe if the method outlined in the section: A method for calculating a regression equation (see page 129), would result in a usable regression equation. Before proceeding with the calculations for the regression equation, first a test for homogeneity was performed. If the result of this calculation was non-significant; the calculation could continue.

An F value for the analysis of co-variance was calculated. The F value related the average treatment sum of squares to the average sum of squares of errors. If F had been non-significant, then it would have been doubtful if any strong relationships could have been expressed in the regression equation. In which case it would not have been worthwhile to proceed with calculating a regression equation with these particular co-variates. Using matrices, an F value of 32.00 was found for the co-variates Categorical-Class and Attitude. this value is significant to $P<0.001$. Hence, hidden within the co-variates Attitude and Categorical-Class there is a significant relationship between the criterion (achievement score) and the co-variates (At and CC).

Yij represented the general regression equation applicable to both experimental and control groups. The Y, as the criterion in the regression equation, represented the achievement score a student with co-variates X1, here Categorical-Class, and X2, here Attitude, was theoretically likely to achieve. The following regression equation was determined:

$$Y_{ij} = -0.14*X1 + 0.12*X2 + 1.46$$

To test the accuracy of the equation the average Categorical-Class and Attitude scores of the experimental and control groups were substituted in the equation.
Experimental : CCav.=2.33, Atav.=41.10: Yav.=6.07
Control : CCav.=3.27, Atav.=38.71: Yav.=5.65
Actual values : Yav.exp=4.92, Yav.control=4.69
Exp % error = (6.07-4.92)/4.92= 23.37%
Control % error= (5.65-4.69)/4.69= 20.47%

Using similar calculations regression equations were calculated for the experimental and control groups separately:

$$Y_{ie} = 0.3801*X1 + 0.0642*X2 - 2.4319$$

Experimental : CCav.=2.33, Atav.=41.10: Yav.=1.09

$$Y_{ic} = -0.32*X1 - 0.05*X2 + 8.58$$

Control : CCav.=3.27, Atav.=38.71: Yav.=5.598

The percentage errors of the individual regression equations are again high. However, the signs (positive or negative) of the coefficients are consistent with the signs of the correlation data found previously.

Two regression equations with co-variates Whole, Categorical-Functional, Categorical-Class, Attitude, Questioning-Recall, Principles-Application

The development of the three regression equations in the previous section showed the method of calculating the regression equations achieved realistic results. However, even though the F value indicated significant relationships between criterion and co-variates, the percentage errors calculated were too large for the regression equations to have of any predictive powers. To reduce these errors, in this section a second set of regression equations was developed to overcome the limitations of just relying on the two co-variates used previously: Categorical-Class and Attitude. As co-variates in this second series the following instrument categories were selected: Whole, Categorical-Class, Categorical-Functional, Attitude, Questioning-Recall and Principles-Application.

Again, as before, first an F value was calculated to find out if there were any significant relationships hidden among the criterion and co-variates. Following along the lines of the trial run, the F value determined using matrices was highly significant to $P<0.001$.

Although the method of calculating were the same as in the trial run, the degree of complexity was increased dramatically due to the presence of six co-variates instead of just two. The general regression equation, Yij, that resulted was:

$$Y_{ij}= -0.2630*W -0.0908*CF -.03009*CC +0.1268*At +0.0057*QR +.0237*PA +2.4074$$

Experimental : W=4.636, CF=4.318, CC=3.273,
At=38.71, QR=-2.96, PA=-1.86: Y=4.66

Control : W=5.62, CF=2.33, CC=2.33,
At=41.10, QR=-1.59, PA=0.32: Y=5.23

Actual values: Y average experimental=4.92, Yaverage control=4.69

Exp % error = (4.66-4.92)/4.92=-5.28 %

Control % error= (5.23-4.69)/4.69= 11.51%

The errors have been considerably reduced compared to the trial run: the errors have been halved. As for the trial, it must be remembered that the F value indicated a significant relationship between criterion and co-variates.

A regression equation applicable only to the experimental group was also calculated:

$$Y_{ie}= 0.2541*W +0.2223*CF +0.3688*CC -0.0738*At -0.0004*QR -0.0006*PA +2.106$$

Experimental : W=4.636, CF=4.318, CC=3.273,
At=38.71, QR=-2.96, PA=-1.86; Y=2.60

Actual values: Y average experimental=4.92

Exp % error = (2.60-4.92)/4.92=-47.23%

The percentage error in Yie is unacceptably high, but it can be seen that at least the signs of the coefficients in the regression equation agree well with the correlation data between the achievement scores and the instrument categories. It appears that the number of co-variables used has exceeded a possible optimum number and the error has increased rather than decreased.

The regression equations Yij and Yie were applied to all treatment classes and the results are displayed in tables 5.23, 5.24 and 5.25.

Table 5.23 Average values of the co-variates for all treatment classes and the control class.

	B.B.	EXP.	GR.MC.	C.M.	ADV.	CONTROL
Whole	4.70	4.64	3.53	4.20	4.42	5.62
Categorical-Functional	4.53	4.32	4.14	4.14	4.17	2.33
Categorical-Class	2.98	3.27	3.30	2.77	4.13	2.33
Attitude	39.16	38.17	35.91	35.19	39.88	41.10
Questioning-Recall	-2.23	-2.96	-5.82	2.25	-0.89	-1.59
Principles-Application	1.69	-1.86	0.89	2.15	-4.06	0.32

Legend: B.B.=Bubbles and Bangs (grade 9)
EXP.=Experimental class (grade 9)
GR.MC.=Green Machine (grade 10)
C.M.=Chemistry in the Market place (grade 10)
ADV.=Advanced Science (grade 10)
CONTROL=Control class (grade 9)

Table 5.24 Predicted criterion score, Y_{ij}, as calculated using the six co-variate general regression equation compared with actual criterion scores for each of the treatment and control classes.

Class	Yij	Yactual	%error rel.	%error abs.
B.B.	4.86	3.98	22.11	9.80
EXP.	4.66	4.92	- 5.28	- 3.30
GR.MC.	4.65	2.24	107.75	30.10
C.M.	4.62	4.64	- 0.47	- 0.30
ADV.	4.58	3.84	19.25	9.30
CONTROL	5.23	3.92	11.51	6.80
average all classes				8.70
av. excl. GR.MC.			8.62	4.46

Legend: Y_{ij}=predicted criterion score calculated from general regression equation
Yactual=actual criterion score
%error rel.=percentage error relative to actual criterion score
%error abs.=percentage error relative to maximum criterion score

B.B.=Bubbles and Bangs (grade 9)
EXP.=Experimental class (grade 9)
GR.MC.=Green Machine (grade 10)
C.M.=Chemistry in the Market place (grade 10)
ADV.=Advanced Science (grade 10)
CONTROL=Control class (grade 9)
av.=average error

Table 5.25 Predicted criterion score, Y_{ie}, as calculated using the six co-variate regression equation (applicable only to treatment classes) compared with actual criterion scores for each of the treatment classes.

Class	Yie	Yactual	%error rel.	%error abs.
B.B.	3.63	3.98	- 8.78	- 4.38
EXP.	2.60	4.92	-47.22	-29.00
GR.MC.	2.49	2.24	11.25	3.13
C.M.	2.52	4.64	-45.78	-26.50
ADV.	2.74	3.84	-28.67	-13.75
average	2.80	3.92	-28.64	-14.10
Control	1.88	4.69	-60.00 (for comparison only)	

Legend: Y_{ie}=predicted criterion score calculated from experimental regression equation
Yactual=actual criterion score
%error rel.=percentage error relative to actual criterion score
%error abs.=percentage error relative to maximum criterion score
B.B.= Bubbles and Bangs (grade 9)
EXP.= experimental class (grade 9)
GR.MC.= Green machine (grade 10)
C.M.= Chemistry in the Market place (grade 10)
ADV.= Advanced Science (grade 10)
average= average for treatment classes
Control= Control class (grade 9)

<u>One regression equation with co-variates Whole, Categorical-Functional, Categorical-Class, Attitude, Enjoyment, Principles, Application, Questioning and Recall</u>

The relatively small contribution of Questioning-Recall and Principles-

Application suggested that a more accurate regression equation could be built by deleting Questioning-Recall and Principles-Application as combined scores and using Questioning, Recall, Principles and Application as separate co-variates. All other co-variates were maintained and another was included as well: Enjoyment. In all, nine co-variates were used.

The increasing number of co-variates made the complexity of the calculations many factors more difficult due to the more than proportional increase in the number of cross-products. The result of these calculations was the formulation of a general regression equation, Y_{ij}:

$$Yij = -0.0714*W -0.3113*CF +0.0757*CC +0.0218*At -0.0083*E +0.0050*P +0.0047*A +0.0013*Q +0.1137*R -0.6407$$

To determine the accuracy of this sixth regression equation tables similar to those set up for previous regression equations were constructed.

Table 5.26 For all treatment, experiment and control classes the values of the co-variates, the predicted criterion score based on the regression equation, the actual criterion score of the class and the percentage error relative to the actual score are displayed.

	EXP.	CON.	B.B.	GR.MC.	C.M.	ADV.
Whole (C.S.C.B.)	4.64	5.62	4.67	3.53	4.20	4.42
Categorical-Functional (C.S.C.B.)	4.32	2.33	4.53	4.14	4.14	4.17
Categorical-Class (C.S.C.B.)	3.27	2.33	2.98	3.30	2.77	4.13
Attitude (T.O.S.R.A.)	38.71	41.10	39.16	35.91	35.91	39.88
Enjoyment (T.O.S.R.A.)	32.33	29.14	33.00	36.13	34.14	39.24
Principles (C.C.P.I.)	49.91	53.45	52.00	51.26	51.25	47.33
Application (C.C.P.I.)	51.77	53.14	49.92	50.37	49.20	52.50
Questioning (C.C.P.I.)	47.68	45.77	47.92	46.00	50.90	47.44
Recall (C.C.P.I.)	50.64	47.36	50.15	52.37	48.15	49.94
Yij	4.82	5.03	4.68	4.58	4.49	4.83
Yactual	4.92	4.69	3.98	2.24	4.64	3.84
%error rel.	-2.0	7.1	17.61	104.33	-3.17	25.65

From the table for the nine co-variate regression equation it can be seen that with this equation the error has now been reduced to 2% for the experimental group and 7% for the control group. This compares to 5% and 11% respectively, for the six co-variate regression equation. Just like the improvement from the two co-variate to the six co-variate regression equation the error has again been halved.

Even though the F values for the analysis of co-variance indicated strong, significant relationships between the criterion and the co-variates, the large errors (relative to actual criterion scores and relative to the maximum possible score) of the two and six co-variate regression equations necessitated the construction of the nine co-variate regression equation.

As well as the relative and absolute errors, another way of checking the accuracy of the regression equation was to calculate the correlation between predicted and actual criterion scores. These correlation calculations were done along the lines suggested by Cohen (1976, 332):

six co-variables	Y_{ij} :	rho= 0.088	non-significant
	Y_{ie} :	rho= 0.297	$P<0.18$
nine co-variables	Y_{ij} :	rho= 0.409	$p<0.15$

It is interesting to note that even though the error for Y_{ie} was greater than the error for Y_{ij} the correlation value rho is much better for Y_{ie} than for Y_{ij}. It could be postulated that if a nine co-variate Y_{ie} regression equation were constructed an even better rho value might result.

Summary of the accuracy data of the six regression equations

Six regression equations were calculated. Some were applicable to all students in both treatment and control groups; these equations were designated: Y_{ij}. Others were applicable only to treatment classes, including the experimental class; these equations were designated: Y_{ie}. The accuracy of these equations was determined in three ways:

(1) calculate the errors both relative to the actual average criterion for a particular group and relative to the maximum possible criterion score (the difference between predicted and actual score divided by eight, expressed as a percentage). The former is called the relative error, the latter is called the absolute error;
(2) calculate the correlation between the predicted criterion score for a particular class and the class' actual criterion score. The significance of this rho value was also calculated;
(3) the Y_{ie} regression equation was applied to the data of the control class. The equation was not intended for this group and ought to show a larger error than the treatment groups for which the equation was intended.

Table 5.27 The six regression equations that predict criterion scores, their relative and absolute errors and their correlation to actual scores.

Regression equation	%error rel.	%error abs.	rho
Yij= -0.14*CC +0.12*At +1.46	exp=23 control=20	exp=14 control=7	n.c.
Yie= 0.3801*CC +0.0642*At -2.4319	exp=-78	exp=-48	n.c.
Yic= -0.32*CC -0.05*At +8.58	control=19	control=11	
Yij= -0.2630*W -0.0908*CF -0.3009*CC +.1268*At +.0057*QR +.0237*PA +2.4074	exp=-5 control=11	exp=-3 control=7	0.09
Yie= 0.2541*W +0.2223*CF +0.3688*CC -.0738*At -.0004*QR -.0006*PA +2.106	exp=-47 control=-60	exp=-29 control=-36	0.30
Yij=-0.0714*W -0.3113*CF +0.0757*CC +0.0218*At -0.0083*E +0.0050*P +0.0047*A +0.0013*Q +0.1137*R -0.6407	exp=-2 control=7	exp=0 control=0	0.41

Legend: n.c.= not calculated.

An improvement that could be made would be to calculate separate regression equations for unsuccessful and successful experimental students. The resulting equations would be applicable only to unsuccessful and successful treatment groups, respectively. Another improvement could be made to the regression equations applicable to experimental students. The number of co-variables could be expanded from six to nine. The improvements in the error of the predicted scores when the two co-variate regression equation was expanded to six co-variates suggests the error in a nine co-variate regression equation would be reduced further. Obviously, a tailor made computer program would be ideal if further improvements were attempted. The time consuming arithmetic is due to two factors:

(1) the increasing complexity of co-variate interaction, for nine co-variates there are 35 cross-products;
(2) the calculations of transposes and inverses of 9x9 matrices using a hand held calculator.

Despite the low correlation values, it should be noted that the errors of the final regression equation are quite good, especially if seen in relation to the fore going regression equations. Also, the F value pertaining to the relationships between criterion and co-variates was determined to be highly significant.

Further improvements will be discussed in chapter 6: **conclusions**. In this chapter the differences between Y_{ie} and Y_{ij} are further explored and another regression equation is constructed, this new regression equation applies only to students in the S.O.L.O. levels Relational, Relational-General Abstract and General Abstract. This equation is not included here because the justification for the choice of its three co-variates is part of the development of a model that shows the student characteristics relevant to coping successfully with the inquiry method.

Having stated the details of the parametric and non-parametric analyses and having presented the results of the statistical calculations, the task of the next chapter is to interpret these results. The conclusions drawn from the results will, where appropriate, refer to information found in the literature review, address the categories of the instruments described in chapter three, find answers to all the research hypotheses detailed in chapter four and list some implications for science teachers.

CHAPTER 6

CONCLUSIONS

6.1 A description of the most important findings

6.1.1. Hypothesis 1

Hypothesis 1(a)

Hypothesis 1(a) was formulated as: in the inquiry method of teaching science successful students have significantly different characteristics than unsuccessful students.

The data collected on the Combined Cognitive Preference Inventory and Cognitive Style of Classification Behaviour instruments clearly showed that successful and unsuccessful students, regardless of their achievement scores, placed emphasis on different characteristics to a significant degree.

Table 6.1 below presents a fraction of the evidence supporting hypothesis 1(a). The table was derived from the strengths of relationships expressed in tables 5.7 (S.O.L.O. versus C.C.P.I..) and 5.8 (S.O.L.O. versus C.S.C.B.). Above the mean student scores were used (table 5.6) with tables 5.10 and 5.11 providing information whether or not successful inquiry students outperformed unsuccessful inquiry students.

Table 6.1 Selected data in support for hypothesis 1(a).

Instrument category	S.O.L.O. level	successful student performance relative to unsuccessful students
Principles	S.O.L.O. Multiple	S scored lower than U.
Application	S.O.L.O. Multiple	S scored higher than U.
Questioning-Recall	S.O.L.O. Relational	S scored higher than U.
Principles-Application	S.O.L.O. Multiple	S scored lower than U.
Principles-Application	S.O.L.O. Relational	S scored higher than U.
Part Whole		S scored lower than U.
Whole		S scored lower than U.
Categorical-Functional		S scored higher than U.
Categorical-Class		S scored higher than U.
Categorical-Inferential		S scored higher than U.

The justification for the acceptance of hypothesis 1(a) is best stated by displaying a model of student characteristics necessary for success in the inquiry method of teaching science. The development of the model is explained in sections: *the correlations between instrument categories and between instrument categories and end of quarter achievement scores* (see page 200) and *Hypothesis 2* (see page 178) of this chapter.

The model, see table 6.2, shows the linear development of different student characteristics. Starting from the left, at S.O.L.O. Singular level students tend to interpret a stimulus in terms of some of its parts. Before focussing on the Application of the stimulus, concrete students have to learn to view the stimulus as a whole. S.O.L.O. Relational and Relational-General Abstract students, also called transitional students, must learn to classify a stimulus in terms of function. At this level, the Principle behind the stimulus and how the stimulus relates to other stimuli must not be emphasized.

Table 6.2 The final 'hierarchy of instrument categories'-model for inquiry students.

Part Whole--> Whole--> A&CF--------> CC----------> P-------> R-------> QR

Part Whole	Whole	A&CF	CC	P	R	QR
		low Principles low Recall	Cat-Functional low Principles low Recall	with low Appl-ication	high Princ-iples	reduce Recall

	CONCRETE	TRANSITIONAL	FORMAL
S.O.L.O. Singular	S.O.L.O. Multiple	S.O.L.O. Relational and Relational-General Abstract	S.O.L.O. General Abstract

Legend: A&CF=Application and Categorical-Functional
CC=Categorical-Class
P=Principles
R=Recall
QR=Questioning-Recall

Only students at the S.O.L.O. General Abstract, formal, level can be exposed to the Principles on which the stimulus operates. This change of emphasis must be accompanied by a de-emphasis of Application. Once the Principles of the stimuli are understood, the Recall with other stimuli can be emphasized on the basis of the Principles on which the stimuli rest. Only once these characteristics have been mastered can Inquiry (Questioning-Recall) into stimuli begin.

This model, based on significant statistical data (Mann-Whitney U test, Chi-square, Guttman's lambda, Spearman's rho, Kendall's W, the W.N.P.D. calculations and the inter-instrument categories' correlations), is as important as it is significant. It is important in that it gives teachers a guide on what characteristics to emphasize in order to increase the chances of student success in the inquiry method. The model is also important in that no researchers have ever reported a sequential development of the characteristics measured by the C.C.P.I.. and T.O.S.R.A. instruments; it has always been assumed that students tend to belong predominantly to one particular trend. The literature on the C.S.C.B. instrument has reported a shift away from wholistic descriptors toward more analytical classifications as students matured (chapter 3, page 43). This shift is confirmed by the model.

The model is significant because it provides a basis for improving the likelihood of success with the inquiry method. The introduction in the Research Design chapter argued that the inquiry method was worthwhile and desirable. Therefore, if more students can succeed with this method there is a good argument for including the method more often in the curriculum.

Based on the results of this research, hypothesis 1(a) can be accepted: certain student characteristics can be developed or concentrated upon in order to improve the likelihood of success in the inquiry method of teaching science.

Hypothesis 1(a) details

This hypothesis can be addressed by investigating the model developed in the section: *the correlations calculations between instrument categories and between instrument categories and and of quarter achievement scores* (see page 200). As stated in that section, the model is supported by the conclusions from comparisons among the five treatment classes as well as by the conclusions from the comparisons between experimental and control classes. In a broad sense the model also supports the research conclusions presented in section 3.2.3 (refer page 53) with regard to grade (the higher the grade the more Principles is preferred), achievement and cognitive ability (the higher the achievement score the greater the preference for Principles and high achievers tend to have high QR scores).

Part
Whole->Whole->Application&->Categorical-Class->Principles->Recall->QR
Cat-Functional

S.O.L.O. Singular	S.O.L.O. Multiple	S.O.L.O. Relational	S.O.L.O. General Abstract

Legend: Cat=Categorical
QR=Questioning-Recall (also identified as Inquiry)

Success in inquiry for concrete students can be promoted by using both Part-Whole and Whole to focus on the application and use of objects. Beyond this step it would be advantages to concentrate on classifying items by Utility (Application and Categorical-Functional) and Whole. For Formal students classifying (Categorical-Functional) could be done by using the Principles that underlie the objects or phenomena. Only when Principles of separate objects or phenomena is understood or mastered can the objects be memorised (Recall) and Principles be examined. Project and Free inquiry (see page 33) can be employed to promote Inquiry (Questioning-Recall) and then only once Recall has been mastered.

The model can be further refined by using the results of the non-parametric statistical analysis' steps involving the W.N.P.D. and percentage likelihood calculations. These calculations confirm that Principles and Recall apply mainly to formal students and Application and Categorical-Functional mainly to concrete students. It is, therefore, advisable to keep Principles and Recall low for concrete students, to keep Recall low for transitional students but increase Principles at the expense of Application. For formal students it would be best to maintain Recall at a high level and to keep Principles higher than Application. Only once Recall has been mastered can Recall be reduced and Questioning be increased. At the low end of the cognitive developmental stages, S.O.L.O. Singular and S.O.L.O. Multiple, it appears important to replace Whole as quickly as possible by Categorical-Functional, or Application. Only after Categorical-Functional has been grasped can transitional students spend a lot of time on Categorical-Class, but Application must be kept high and Principles low for these students.

For scientific Attitude there is no significant pattern available in line with some of Fraser's data (1978,513). However, there is sufficient evidence to suggest that time can be spent developing Attitude but not at the expense of Application; it appears that Attitude is mainly appropriate to transitional and formal students. Enjoyment of science appears fairly important for all students, even at the expense of Attitude! However, Categorical-Class is significantly more important than Enjoyment! This is important, because this indicates that science activities should be selected more on the basis of Categorical-Functional and Categorical-Class than on the basis of student enjoyment and, perhaps, even interest.

These further refinements can be inserted in the model as follows:

```
Part
Whole->Whole->Application&---> Cat-Class----> P-------> R---------> QR
           |   Cat-Functional| with
           |   low Principles |Cat-Functional  low Appl- high Princ- reduce
           |   low Recall     |low Principles  ication   iples       Recall
           |                  |low Recall     |
-------------------------------------------------------------------------
           |   CONCRETE       |TRANSITIONAL   |  FORMAL
-------------------------------------------------------------------------
S.O.L.O.   |   S.O.L.O.       |S.O.L.O.       |  S.O.L.O.
Singular   |   Multiple       |Relational     |  General Abstract
```

Legend: Cat=Categorical
P=Principles
R=Recall
QR=Questioning-Recall (also known as Inquiry)
Note: this model is refined further later in the chapter.

Although a controlled experiment would have to be conducted to verify the suggestions made on the basis of the model, based on the results of this research hypothesis 1(a) can be accepted: certain student characteristics can be developed or concentrated upon in order to improve the chances of success in the inquiry method of teaching science. It should be noted that in the literature on the C.S.C.B. instrument (see page 43), it had been reported that as students matured there was a shift away from Wholistic descriptors to more analytical categorisation. This fits in very well with the model. Also, it was noted that the use of Categorical-Functional labelling decreased with age and this too agrees with the model.

Hypothesis 1(b)

Hypothesis 1(b) was formulated as: the difference between successful and unsuccessful students is not the same for S.O.L.O. Singular and Multiple students as for S.O.L.O. Relational and General Abstract students.

Due to the small number of S.O.L.O. Singular and General Abstract students to achieve statistical significance a comparison could only be made between S.O.L.O. Multiple and Relational students. Tables 6.3 and 6.4 show those instrument categories which had a significant difference between successful and unsuccessful students. These tables used some of the data in table 6.1, but present the information in a different way. Further data have been taken from tables 5.7 and 5.8 in order to present a breakdown by S.O.L.O. level.

Table 6.3 Details of the differences between successful and unsuccessful students on the Combined Cognitive Preference Inventory instrument.

Instrument category	details of the differences between successful and unsuccessful inquiry students
Principles	S.O.L.O. Multiple: successful scored lower than unsuccessful
	S.O.L.O. Relational: no significant difference
Application	S.O.L.O. Multiple: successful scored higher than unsuccessful
	S.O.L.O. Relational: no significant difference
Questioning -Recall	S.O.L.O. Multiple: no significant difference
	S.O.L.O. Relational: successful scored higher than unsuccessful
Principles-Application	S.O.L.O. Multiple: successful scored lower than unsuccessful
	S.O.L.O. Relational: successful scored higher than unsuccessful

Table 6.4 Details of the differences between successful and unsuccessful students on the Cognitive Style of Classification Behaviour instrument.

Instrument category	details of the differences between successful and unsuccessful inquiry students
Part Whole	S.O.L.O. Multiple: successful scored lower than unsuccessful; S.O.L.O. Relational: successful scored lower than unsuccessful; difference between successful an unsuccessful less for S.O.L.O. Relational than for Multiple
Whole	S.O.L.O. Multiple: successful scored lower than unsuccessful; S.O.L.O. Relational: successful scored lower than unsuccessful; both S.O.L.O. Relational successful and unsuccessful scored higher than S.O.L.O. Multiple successful and unsuccessful
Categorical -Functional	S.O.L.O. Multiple: successful scored higher than unsuccessful S.O.L.O. Relational: successful scored higher than unsuccessful; both S.O.L.O. Relational successful and unsuccessful scored lower than S.O.L.O. Multiple successful and unsuccessful
Categorical -Class	S.O.L.O. Multiple: successful scored higher than S.O.L.O. Relational: successful scored higher than unsuccessful; the difference between successful and unsuccessful was greater for S.O.L.O. Relational than for S.O.L.O. Multiple
Categorical -Inferential	S.O.L.O. Multiple: successful scored higher than unsuccessful; S.O.L.O. Relational: successful scored higher than unsuccessful; both S.O.L.O. Relational successful and unsuccessful scored higher than S.O.L.O. Multiple successful and unsuccessful; the difference between successful and unsuccessful was greater for S.O.L.O. Relational than for S.O.L.O. Multiple

Tables 6.3 and 6.4 provide the justification for the acceptance of hypothesis 1(b).

Hypothesis 1(c)

Hypothesis 1(c) was formulated as: students studying Chemistry, Physics and Biology by the inquiry method have significantly different characteristics for each of the disciplines.

There was no conclusive evidence to indicate whether the developmental stage of the student was more important than the topic the student was studying, or vice versa. What can be stated with certainty is that the difference in characteristics between successful and unsuccessful students was more related to developmental stage than to the science topic. In this research the science topics fell in to the broad categories of Chemistry, Physics and Biology. For both grade nine and ten, data on student characteristics showed that Biology (Green Machine) was different from both Chemistry and Physics. The latter two were similar on Categorical-Functional, Categorical-Class and Questioning-Recall in their correlations with achievement scores. Biology appeared to be the odd one out. On inter-instrument correlations Biology differed again from Chemistry and Physics on Categorical-Class versus Categorical-Functional, Principles-Application versus Questioning-Recall and Principles-Application versus Attitude. Also, Biology was the one class that followed the developed regression equations least well.

Therefore, Chemistry and Physics were similar and both were different compared with Biology. Biology differed on the bases of student success characteristics, inter-instrument correlations and adherence to regression equations.

Hypothesis 1(c) is not supported on the basis of lack of statistical significance.

6.1.2 Hypothesis 2

Hypothesis 2(a)

Hypothesis 2(a) was formulated as: the inquiry method of teaching science produces end of quarter achievement scores equal to or better than the conventional science teaching method.

The data in tables 6.5 and 6.6 show that hypothesis 2(a) can be accepted on criteria Pr-1, C-1, I-3 and K for S.O.L.O. Singular and Multiple students and on all criteria for S.O.L.O. Relational and General Abstract students. Table

6.5 shows a summary of the performance of the experimental group compared to the control group. This table is a much refined form of table 5.4. Rather than a class by class grouping, students have been broken down into S.O.L.O. levels. More details of the assessment criteria are provided in table 6.5 compared with table 5.4 which only has the average achievement score of selected criteria. Table 6.6, in abbreviated form, is shown the content of each of the criteria mentioned in table 6.5.

Table 6.5 Summary of a comparison between experimental, Te, and control, Tc, classes on the basis of achievement scores.

S.O.L.O. level	achievement score comparison	achievement score criteria	significance
Relational & General Abstract	experimental > control	all criteria	$p<0.08$
Relational	experimental > control	Pr-1,C-1,I-3,K	$p<0.06$
Relational	experimental = control	Pr-2,Pr-3,I-4,I-1	
Multiple	experimental > control	Pr-1,C-1,I-3,K	$p<0.09$
Multiple	experimental < control	Pr-2,Pr-3,I-4,I-1	$p<0.09$
all levels	experimental > control	all criteria	$p<0.09$
all levels	experimental > control	Pr-1,C-1,I-3,K	$p<0.09$
all levels	experimental < control	Pr-2,Pr-3	$p<0.08$
		I-1,I-4	$p<0.09$

Table 6.6 Eight criteria of the Tasmanian Certificate of Education on which inquiry students differed from control students.

Criterion	content of each criterion
Pr-1	making qualitative observations and recognizing patterns and trends.
C-1	extracting information independently.
I-3	explaining events and results.
K	using scientific language to express ideas and explanations.
Pr-2	use of apparatus/equipment.
Pr-3	using apparatus accurately.
I-4	using tables, graphs, etcetera, to communicate and make predictions.
I-1	re-arranging raw data into different presentations, for example tables, graphs, etcetera.

Hypothesis 2(a) details

Formal students in the treatment groups performed significantly better on all achievement score criteria compared to the control group. Transitional students performed significantly better on Pr-1, C-1, I-3 and K, and performed equally well on all other criteria. S.O.L.O. Multiple inquiry students performed significantly better on Pr-1, C-1, I-3 and K, but performed worse (although not all at a significant level) on Pr-2, Pr-3, I-4 and I-1.

A concept is the combination of a model and the verbalisation of that model. Therefore, observations and explanations in appropriate scientific language (Pr-1, I-3 and K) indicate that the quality of learning is higher for the inquiry method than for conventional methods.

Inquiry students are not necessarily forced to use specific apparatus and are not formally acquainted with their use. Therefore, Pr-2 and Pr-3 scores were not as good for inquiry students as for conventional students. Inquiry students were also not directed at presenting their data in any particular way, although they were encouraged to present data in alternative ways. Apparently, low ability (S.O.L.O. Singular and Multiple) students did not have the necessary logic structures to be able to suggest different data presentations, nor were they able to draw conclusions from them (I-1 was only marginally worse for inquiry students, but I-4 was significantly worse for S.O.L.O. Multiple inquiry students). These data suggest that in conventional science teaching methods the teacher did most of the work to get students to draw up tables in different ways and draw conclusions from the re-arranged data. That is, it appears doubtful that students would have been able to do these particular two techniques by themselves afterwards and would have been able to apply these techniques to different situations (transfer is unlikely).

In light of these comments, it appeared that even for S.O.L.O. Multiple students inquiry achieved more meaningful learning than conventional teaching methods. In future inquiry courses S.O.L.O. Multiple students need more help with analysing their data from a logical/mathematical point of view. This suggestion is reinforced by the model in that S.O.L.O. Multiple students see a stimulus as a whole (Whole) in Pr-1. They can explain, presumably, the results of their experiments only in terms of what happened (Categorical-Functional) and they are incapable of abstraction or generalizing in I-3. Also, they are unable to extract the principles (Principles) nor remember the results and conclusions (Recall) of their experiments in I-4.

Hypothesis 2(b)

Hypothesis 2(b) was formulated as: regression equations for the inquiry and conventional science teaching methods will be significantly different.

In total seven regression equations were developed from the data comparing the experimental and control classes. The regression equations are presented in table 6.7. The capital letter Y refers to the predicted student achievement score (criterion), the subscript i refers to individual student (as opposed to class average) and the subscripts j, e and c refer to all students, experimental students and control students, respectively.

Table 6.7 Collation of all regression equations calculated.

Two co-variables:

Y_{ij}= -0.14*CC +0.12*At +1.46

Y_{ie}= 0.38*CC +0.06*At -2.43

Y_{ic}= -0.32*CC -0.05*At +8.58

Six co-variables:

Y_{ij}= -0.26*W -0.09*CF -0.30*CC +0.13*At +0.01*QR +0.02*PA +2.41

Y_{ie}= 0.25*W +0.22*CF +0.37*CC -0.07*At -0.001QR -0.002PA +2.11

Nine co-variables:

Y_{ij}= -0.07W -0.31CF +0.08CC +0.02At -0.001E
+0.002P+0.003A +0.001Q +0.11R -0.64

(continued over page)

Three co-variables for S.O.L.O. Relational, Relational-General Abstract and General Abstract:
Y_{ie}= 0.54*CC -0.49*P -1.00*R +119.77

Legend: W=whole (C.S.C.B.)
CF=Categorical-Functional (C.S.C.B.)
CC=Categorical-Class (C.S.C.B.)
At=Attitude (T.O.S.R.A.)
E=Enjoyment (T.O.S.R.A.)
QR=Q-R=Questioning-Recall (also known as Inquiry) (C.C.P.I..)
PA=P-A=Principles-Application (also known as Utility) (C.C.P.I..)
P=Principles (C.C.P.I..)
A=Application (C.C.P.I..)
Q=Questioning (C.C.P.I..)
R=Recall (C.C.P.I..)

Hypothesis 2(b) was accepted for the six co-variable Y_{ie} accurate to 14% error and a correlation of 0.09 and the three co-variable Y_{ie} (the latest developed) with 1% error and a correlation of -0.22 compared with the nine co-variable Y_{ij} accurate to 9% absolute error and a correlation of 0.41 with actual achievement scores. When the three co-variable Y_{ie} was applied solely to formal students for whom it was designed, the correlation became 1.00 with an error of 2%. In perspective, it should be noticed that in educational and psychology statistics correlations greater than 0.7 are rare and that most of the correlations reported in the literature are mostly less than 0.4 (Nunnally, 1975, 155).

Hypothesis 2(b) details

The conclusions related to this hypothesis and the development of regression equations used the data gathered from the comparison of the experimental and treatment classes, that is from the parametric analysis. At the start of this section a tentative model of the hierarchy of instrument categories is suggested on the basis of inter-correlation data between achievement scores, C.C.P.I.., C.S.C.B. and T.O.S.R.A. instrument categories that were used to develop the model for inquiry students. This model showed that the inquiry method did indeed place specific emphases on different instrument categories.

Table 6.8 Similarities and differences of inter-instrument correlation data between experimental and control classes.

Experimental similar to control	experimental different from control
% vs W,CF,E,QR	% vs CC,At,PA
W vs CF,CC,At,QR,%	W vs E,PA
CF vs CC,At,QR,%,W	CF vs PA
CC vs W,CF	CC vs At,E,QR,PA,%
At vs E,PA,W,CF	At vs QR,%,CC
E vs PA,%	

Legend: %= achievement score
W=whole (C.S.C.B.)
CF=Categorical-Functional (C.S.C.B.)
CC=Categorical-Class (C.S.C.B.)
At=Attitude (T.O.S.R.A.)
E=Enjoyment (T.O.S.R.A.)
QR=Q-R=Questioning-Recall (also known as Inquiry) (C.C.P.I..)
PA=P-A=Principles-Application (also known as Utility) (C.C.P.I..)
P=Principles (C.C.P.I..)
A=Application (C.C.P.I..)
Q=Questioning (C.C.P.I..)
R=Recall (C.C.P.I..)

The following five conclusions could be drawn from the correlation data (refer tables 6.8 and 5.19):

(1) Whole was irrelevant to both experimental and control students with regard to achievement score. This category must have been a prerequisite for conventional students because it was a prerequisite for inquiry students. Unlike the inquiry students, Whole correlated well to Principles for conventional students. The relation between Whole and Categorical-Functional appeared the same for both groups. Judging from the achievement score correlation with Whole, Whole was not a handicap for conventional students;
(2) Categorical-Functional appeared to be of little use for conventional

students, because there was a positive correlation between achievement score and Principles-Application, but a zero correlation between Categorical-Functional and Principles-Application. There was also a zero correlation between Enjoyment and Categorical-Functional, but a positive correlation between Enjoyment and achievement score. Therefore, conventional students emphasized Principles.

(3) for conventional students, Categorical-Functional and Categorical-Class appeared to be totally different skills and there appeared to be a relation between Recall and Categorical-Functional. In fact, Categorical-Class appeared irrelevant and detrimental to Principles for conventional students.

(4) Attitude was not as strongly correlated to Enjoyment for conventional students (0.40) as it was for inquiry students (0.75). Also, Attitude appeared irrelevant to Questioning-Recall and Principles-Application for conventional students.

(5) unlike inquiry students, for conventional students Enjoyment was very strongly correlated to achievement score. Enjoyment was irrelevant to Categorical-Class and Categorical-Functional, Enjoyment was somewhat correlated to Whole and strongly correlated to Principles.

The result of these five conclusions was the following tentative model applicable to conventional students:

$$\text{Whole}\left\{\begin{array}{l}\dashrightarrow \text{Categorical-Functional} \dashrightarrow \text{Recall} \dashrightarrow \\ \\ \dashrightarrow\end{array}\right\}\text{Principles}$$

Note that there appeared to be alternative paths to Principles. Enjoyment was very important overall. Compared to the model for inquiry students, conventional students did not appear to use many instrument categories. This agreed well with the literature which stated that inquiry students were more fluent and flexible in their use of different styles of categorization than non-inquiry students (see page 43).

The irrelevancy of Categorical-Class for conventional students is illustrated by comparing the regression equations Y_{ie}, the regression equation appropriate to inquiry students, and Y_{ic}, the regression equation appropriate to conventional students, for the instrument categories Categorical-Class and Attitude:

(1) in $Y_{ie}=0.38*CC + 0.06*At - 2.43$: Categorical-Class was a large contributor;
(2) in $Y_{ic}=-0.32*CC - 0.05*At + 8.58$: Categorical-Class contributed negatively.

It appears Attitude was not important and the difference between the coefficients 0.32 and 0.05 showed clearly that Y_{ic} was dependent on an instrument category other than Attitude. The negative sign in front of 2.43 indicated the importance of Categorical-Class was over-estimated, whereas the positive sign in front of 8.58 indicated the detrimental effect of Categorical-Class was over-estimated.

The model for conventional students did indeed look very different than the model for inquiry students. Next, the regression equations for inquiry, conventional and all students (Y_{ie}, Y_{ic} and Y_{ij}, respectively) were compared.

The six co-variate general predictive regression equation to predict the achievement score on the basis of certain instrument categories showed an average error of +8.7% (including the Biology inquiry class), ranging from +10% to -3% with the Biology class the only exception with an error of 30%. Significance calculations (Dayton, 1970) showed a non-significant deviation between predicted and actual achievement scores. That is, the six co-variate general regression equation was acceptable for all science classes regardless of the method by which the students were taught. Unfortunately, the correlation between predicted and actual achievement scores was only 0.09, an unacceptable low value indicating the error was not systematic.

The six co-variate experimental regression equation had an average error of -14%, which was a larger error than the general equation, but when the control class was used for comparison on the experimental regression equation an error of -35% resulted. This showed that despite the large error, the experimental equation was appropriate only to inquiry classes. As for the general equation the deviation between predicted and actual achievement scores was non-significant indicating the inquiry regression equation was acceptable. Unfortunately, a lack of systematic error caused a low (0.09) correlation between predicted and actual scores.

The improved nine co-variate general regression equation had a slightly better error, 8.6%, the F ratio between predicted and actual achievement scores was again non-significant (that is, acceptable) and this time the correlation had improved to 0.41. Therefore, the general nine co-variate regression equation was an acceptable representation of how the selected co-variates interacted with the degree of achievement in science courses regardless of the method by which students were taught. It has to be acknowledged that the errors in the regression equations will not be totally eliminated as long as factors like classroom environment are excluded from the regression equation. For instance, classroom environment has been reported as being responsible for 5 to 14% of variance in achievement in science (Talton and Simpson, 1987, 507). Also, with regard to the number of co-variates included in the equations, Okebukola (1987, 119) reported that twelve factors accounted for only 64% of the variance in achievement.

The relative importance of different instrument categories for general science students was as follows, as based on the general regression equation:

Whole	contributes negatively and was undesirable;
Categorical-Functional	very undesirable, the conventional teaching method

	did not stress Categorical-Functional;
Categorical-Class	very useful, indicating students were required to classify a lot of the time;
Attitude	third most important category of all categories;
Enjoyment and	
Principles-Application	both contributed little;
Questioning	did not contribute at all;
Recall	very useful, second most important, indicating students were encouraged to memorize the results and conclusions of their experiments after classifying phenomena or objects.

To visualize the different emphases the regression equations placed on different instrument categories, for the inquiry classes compared to general science classes, it is helpful to compare the relative weighting of each of the instrument categories in the three most promising regression equations. This is done in table 6.9 below.

Table 6.9 The relative weighting of importance of each co-variate with Questioning or Questioning-Recall taken arbitrarily as unity.

Eqn.	W	CF	CC	At	E	P	A	Q	R	QR	PA
Y_{ie}^6	2338	2045	3393	-185	NA	NA	NA	NA	NA	-1	-2
Y_{ij}^6	-170	-59	-194	22	NA	NA	NA	NA	NA	1	6
Y_{ij}^9	-202	-88	214	17	-6	5	5	1	123	NA	NA

Legend: Eqn.=regression equation
Y_{ie}^6=six co-variate experimental regression equation
Y_{ij}^6=six co-variate general regression equation
Y_{ij}^9=nine co-variate general regression equation
W=whole (C.S.C.B.)
CF=Categorical-Functional (C.S.C.B.)
(continued over page)
CC=Categorical-Class (C.S.C.B.)
At=Attitude (T.O.S.R.A.)

E=Enjoyment (T.O.S.R.A.)
QR=Q-R=Questioning-Recall (also known as Inquiry) (C.C.P.I..)
PA=P-A=Principles-Application (also known as Utility) (C.C.P.I..)
P=Principles (C.C.P.I..)
A=Application (C.C.P.I..)
Q=Questioning (C.C.P.I..)
R=Recall (C.C.P.I..)
NA=not applicable

While the data in the table speak largely for themselves, it is interesting to note the similarity of weightings between $Y_{ij}{}^6$ and $Y_{ij}{}^9$. The only surprise between these two equations was the reversal of importance of Categorical-Class from -194 to +214. It seems very likely this reversal was due to the unexpected importance of Recall. Unexpected because Questioning-Recall in $Y_{ij}{}^6$ is simply +1 whereas a negative sign would have been expected to indicate a negative correlation between achievement score and Questioning and, therefore, a positive correlation between achievement score and Recall.

Overall, the conclusion is that Y_{ie} was very different from Y_{ij}, indicating that inquiry did require different student characteristics or emphasized different instrument categories than conventional methods of teaching science:

Y_{ij} emphasizes	Categorical-Class and Recall, avoids Categorical-Functional and Whole and makes Principles, Application and Questioning irrelevant;
Y_{ie} emphasizes	Categorical-Class, Categorical-Functional and Whole, avoids Attitude and makes irrelevant Principles-Application and Questioning-Recall.

Hence, hypothesis 2(b) was accepted.

The rest of this section is devoted to the development of a simpler regression equation based totally on the hierarchy of instrument categories model for inquiry students. A good model should not only be easily understood and have an application, it should also be able to make predictions from which fluid inquiry can emerge. Thus, to test the model a new regression equation was proposed.

From the inquiry model of category sequences, it should be noted that Categorical-Functional and Whole were concrete skills and Categorical-Class was a transitional skill. It can also be seen that Y_{ie} was most appropriate to classes with mainly concrete students judging by the emphasis on concrete specific instrument categories. Therefore, a regression equation for successful S.O.L.O. Relational and General Abstract inquiry students should emphasize Categorical-Class, Principles and Recall as co-variates. The standard calculations were performed to derive a regression equation based on successful S.O.L.O. Relational and General Abstract students. The resulting regression equation was derived:

$$Yie=0.54*CC - 0.49*P - 1*R + 119.77$$

To test the accuracy of this regression equation the percentage errors were calculated as before. The results are displayed in table 6.10.

From this new regression equation four conclusions could be made:

(1) both relative and absolute errors have improved;
(2) the correlation with actual scores has much improved compared to $Y_{ie}{}^{6}$, eventhough there were now only three co-variates, from 0.09 to -0.22. The negative coefficient was probably due to the inclusion of S.O.L.O. General Abstract students who emphasize Principles and Recall;
(3) most improvement has been made in classes with a high percentage of S.O.L.O. Relational, Relational-General Abstract and General Abstract students;
(4) biology (Green Machine) appeared not to adhere to the model.

Table 6.10 Errors of the regression equation developed on the basis of the hierarchy of instrument categories model for inquiry students.

Class	Y_{ie}	Yactual	%rel. error	%abs. error	Yie better or worse than Y_{ie}^{6}(%abs. error)
Bubbles&Bangs	4.18	3.98	5.03	2.50	1.88% better
absolute					
experimental	3.69	4.92	-25.00	-15.38	13.62% better
absolute					
Green Machine	4.10	2.24	83.04	23.25	20.12% worse absolute
Chemistry in the					
Market place	4.32	4.64	- 6.90	- 4.00	22.50% better absolute
Advanced Science	3.88	3.84	1.04	0.50	13.25% better absolute
treatment average			11.44	1.37	

correlation Y_{ie} and Yactual=-0.22 for all classes (all S.O.L.O. levels)

correlation Y_{ie} and Yactual=1.00 for CM and AdvSc (S.O.L.O. Relational and General Abstract)

Legend: %rel. error=percentage error relative to actual score
%abs. error=percentage error relative to max score of 8

It may be concluded that the model was correct, particularly for the physical sciences. The sign of the coefficients of Principles and Recall are negative and that of Categorical-Class positive. This indicated that, according to the model, the assumption that S.O.L.O. Relational, Relational-General Abstract and General Abstract students were all formal was incorrect. Thus, only true S.O.L.O General Abstract students were formal and S.O.L.O Relational-General Abstract and Relational students were transitional. Even in terms of helping to explain data the model turns out to be powerful.

Consequently, the model could again be improved as displayed in table 6.11.

Table 6.11 The final hierarchy of instrument categories model for inquiry students.

Part Whole->	Whole->A&CF------------>	CC----------->	P--------->	R-------->	QR
	low Principles low Recall	with Cat-Functional low Principles low Recall	low Appl- ication	high Princ- iples	reduce Recall
	CONCRETE	TRANSITIONAL		FORMAL	
S.O.L.O. Singular	S.O.L.O. Multiple	S.O.L.O. Relational& Relational-General Abstract		S.O.L.O. General Abstract	

Legend: W=whole (C.S.C.B.)
CF=Categorical-Functional (C.S.C.B.)
CC=Categorical-Class (C.S.C.B.)
At=Attitude (T.O.S.R.A.)
E=Enjoyment (T.O.S.R.A.)
QR=Q-R=Questioning-Recall (also known as Inquiry) (C.C.P.I..)
PA=P-A=Principles-Application (also known as Utility) (C.C.P.I.)
P=Principles (C.C.P.I..)
A=Application (C.C.P.I..)
Q=Questioning (C.C.P.I..)
R=Recall (C.C.P.I..)

In summary, Y_{ij}^9 is accurate to 9% absolute error and a correlation of 0.41 with actual achievement scores compared to Y_{ie}^6 accurate to 14% error and a correlation of 0.09 and Y_{ie}^3 (the latest developed) with 1% error and a correlation of -0.22. When Y_{ie}^3 is applied solely to formal students for whom it was designed, the correlation becomes 1.00 with an error of 2%. In perspective, it should be noticed that in educational and psychology statistics correlations greater than 0.7 are rare and that most of the correlations reported in the literature are mostly less than 0.4 (Nunnally, 1975, 155). Hypothesis 2(b) is accepted.

6.2 Summarized findings of the steps in the non-parametric statistical analysis

There were six steps in the non-parametric analysis:

(1) the Mann-Whitney U test;
(2) the Chi-square test;
(3) Guttman's Coefficient of predictability, lambda;
(4) Spearman's Rank Order correlation coefficient and Kendall's coefficient of concordance;
(5) the Weighted Net Percentage Difference (W.N.P.D.) analysis of variance;
(6) inter-instrument correlation calculations between different instrument categories.

Each step is discussed in turn. The parametric analysis is not listed as a separate section because it was discussed in detail in the section: Hypothesis 2(b).

6.2.1 The difference between successful and unsuccessful students in terms of end of quarter achievement scores as calculated by the Mann-Whitney U test

There was a significant difference between unsuccessful and successful students; successful students outscored unsuccessful students. This finding was important because the distinction between successful and unsuccessful was based purely on teacher observation on certain student behaviours but was not based on hard quantitative data. This finding is also important because the difference suggested different regression equations would apply and, therefore, the two categories successful and unsuccessful placed different emphases on different instrument categories. Hence, it appears that certain characteristics could be encouraged in students to increase the probability of success in inquiry science courses. This was probably the single most important application of this study.

6.2.2 The difference between successful and unsuccessful students in terms of C.C.P.I.. and C.S.C.B. instrument categories as calculated by the Chi-square test

The result of the outcome of the calculations of the Chi-square test confirmed the appropriate use of the instruments used to collect data on students. T.O.S.R.A., C.C.P.I.. and C.S.C.B. all could be used to distinguish between unsuccessful and successful students, using above average category scores. Note that this result was different from the conclusion reached in the Mann-Whitney U test, because the conclusion in the Mann-Whitney U test was based on achievement scores, not on instrument categories. The results of the Chi-square test allowed the regression equations to be expressed in terms of the instrument category scores.

6.2.3 The strongest differentiators between successful and unsuccessful students in terms of C.C.P.I.., C.S.C.B. and T.O.S.R.A. categories as determined by Guttman's coefficient of predictability, lambda

Lambda was calculated for each of the C.C.P.I.., C.S.C.B. and T.O.S.R.A. instruments in turn. Therefore, it appears logical to present the findings to each of these instruments separately.

Combined Cognitive Preference Inventory

In general, successful students scored lower in Principles than unsuccessful students, but for S.O.L.O. Relational students this difference did not exist. Therefore, for concrete students, little emphasis should be given to Principles. This did not apply to transitional students. Successful students scored higher in Application than unsuccessful students, especially S.O.L.O. Multiple students. For S.O.L.O. Relational students there was no difference on Application.

Therefore, for concrete students a lot of emphasis should be placed on Application. For transitional and formal students this did not apply.

S.O.L.O. Multiple unsuccessful students scored higher in Questioning-Recall (Inquiry) than successful students. S.O.L.O. Relational unsuccessful students scored much lower in Questioning-Recall than successful students. Therefore, for concrete students an inquiring oriented attitude may be a disadvantage. However, for transitional and formal students an inquiring oriented mind is a definite advantage. With the exception of S.O.L.O. Relational students, unsuccessful students outscored successful students in Principles-Application (Utility), this was because concrete successful students scored much higher in Application and lower in Principles than concrete unsuccessful students. Therefore, for concrete students it is more useful to think in terms of Principles and Application separately than as Principles-Application, where the Application score is subtracted from the Principles score to give one score. Transitional and formal students, unsuccessful and successful, scored approximately the same in Principles and Application. Therefore, the Principles-Application category was a better separator for these students. Also, it may be concluded that for transitional and formal students focussing on utility is an advantage but not as much an advantage a a high Questioning-Recall score.

Cognitive Style of Classification Behaviour

Based on the finding that unsuccessful students tended to have higher Part-whole scores than successful students, Part-whole must be a disadvantage. The same finding applied to the category Whole: Whole was a disadvantage.

The reverse was true for Categorical-Functional scores: unsuccessful students tended to have lower Categorical-Functional scores than successful students, hence Categorical-Functional must be an advantage. High Categorical-Class scores and Categorical-Inferential attributes scores were also a definite advantage because unsuccessful students tended to have much lower Categorical-Class and CI scores than successful students.

Because students with Categorical-Functional, Categorical-Class and Categorical-Inferential styles of categorisation did better in inquiry than students with other styles, perhaps the inquiry method of teaching science promoted these particular styles at the expense of other styles.

Test Of Science Related Attitudes

The conclusions relating to this instrument were encompassed in the inter-instrument categories correlation conclusions. T.O.S.R.A. was not part of the Guttman analysis. For consistency and clarity a general comment can be made here. It can be stated that only two scales from this instrument were found to be useful in this research: Attitude to science and Enjoyment of science. The other scales were found to be insignificant discriminators and did not appear in the regression equations. In the regression equations Attitude was found to be only moderately important (more important than some, less important than others); Enjoyment was found to be only marginally important.

6.2.4 The determination of significant relationships between instrument category scores and end of quarter achievement scores using Spearman's rank order correlation coefficient and Kendall's coefficient of concordance

This fourth step in the statistical analysis related the achievement scores to the instrument categories which were used to express student characteristics. The previous three steps investigated the relationships between the unsuccessful/successful dichotomy on the one hand and achievement scores and instrument categories on the other hand. The fourth step did not word its findings in terms of the unsuccessful/successful dichotomy. In total, nine conclusions,applicable to inquiry students, could be formulated from the results of this step:

(1) consistent with the first conclusion of the section on C.C.P.I.. from step three, the correlation between Principles and achievement score was either zero or negative;
(2) consistent with the second conclusion in the C.C.P.I.. section, the correlation between Application and achievement score was 0.39 for concrete students and zero for non-concrete students;
(3) for grade nine, a -0.35 correlation between Principles-Application and achievement score was found for concrete students. Therefore, for concrete students Application was more important than Principles. For grade ten, this was not the case; a positive correlation between Principles-Application and achievement score was found for all students. Therefore, for Criterion Based Assessment Application was more important than Principles for concrete students. For grade ten, an above average score on the Principles-Application (Utility) scale was important, that is the Principles score must be greater than the Application score. This was consistent with the fourth conclusion in the C.C.P.I.. section: for transitional and formal students an above average Principles-Application score was an advantage;
(4) for students assessed by Criterion Based Assessment in general and for concrete learners in general a high Questioning-Recall (Inquiry) score was a disadvantage. For formal students, but not necessarily for transitional students, an above average Questioning-Recall score was a definite advantage;
(5) consistent with the first conclusion in the C.S.C.B. section, an above average Whole score was a disadvantage. The analysis showed Whole had a negative or zero correlation with achievement score;
(6) for concrete learners, an above average Categorical-Functional score was a disadvantage. Categorical-Functional had a negative or zero correlation with achievement score;
(7) especially for students judged by the Criterion Based Assessment method, in this study grade nine students, an above average Categorical-Class score was an advantage. Categorical-Class had a positive correlation of 0.38, on average, with achievement score;

(8) there was a clear difference between standard assessment method and Criterion Based Assessment: students with an above average Attitude score, reflecting a positive scientific attitude, were at an advantage in C.B.A., but at a disadvantage when standard assessment was used (correlation Attitude with achievement score is -0.40 for standard assessment);
(9) an above average score on the Enjoyment of science scale, E, was an advantage for all students (correlation between Enjoyment and end of quarter achievement score was 0.44).

6.2.5 The relationships between successful-unsuccessful student classification, instrument categories, student S.O.L.O. levels and end of quarter achievement scores as determined by the Weighted Net Percentage Difference (W.N.P.D.) analysis

The results of the W.N.P.D. analysis clearly demonstrated the differences between successful and unsuccessful students on the basis of student characteristics. The data showed successful and unsuccessful students differed on some instrument categories more than others. Also, this difference varied between S.O.L.O. levels (p149, table 5.16). Correlation results of the double comparisons (p151,table 5.17: comparisons 3-7) indicated that the differences between successful and unsuccessful students were similar to the differences between above average achievement score and below average achievement score students. That is, any comparisons between successful and unsuccessful students would have resulted in similar conclusions that would have been reached by comparing above average and below average students.

Therefore, a model, that depicts student characteristics likely to lead to success in the inquiry method, based on successful students should also reflect the characteristics needed to achieve above average achievement scores. This conclusion is used in the next section for the development of such a model.

6.2.6 The correlations between instrument categories and between instrument categories and end of quarter achievement scores

In the regression equations many instrument categories were identified by their abbreviations. At this stage, it may be helpful to list the abbreviations, their proper naming and which instruments the categories belong to before continuing with the conclusions.

Table 6.12 List of instrument categories and their abbreviations.

Abbreviation	instrument category
C.C.P.I..	Combined Cognitive Preference Index
P	Principles
Q	Questioning
R	Recall
A	Applications
PA or P-A	Principles-Application (also known as Utility)
QR or Q-R	Questioning-Recall (also known as Inquiry)
C.S.C.B.	Cognitive Style of Classification Behaviour
P	Part-whole
W	Whole
FI	Functional-Interdependence
CF	Categorical-Functional
CC	Categorical-Class
CI	Categorical-Inferential
T.O.S.R.A.	Test Of Science Related Attitudes
At	Attitude to scientific inquiry
E	Enjoyment of science lessons

An in-depth study of the inter-instrument correlation data revealed an imbedded hierarchy of instrument categories. The hierarchal nature of the instrument categories can be seen in the following three examples.

Example 1: Categorical-Class versus Attitude and Categorical-Class versus achievement score had positive correlations for concrete students; therefore, Categorical-Class and Attitude are desirable for concrete students.

Questioning-Recall versus achievement score had a positive correlation for formal students; therefore, Questioning-Recall is desirable for formal students.

Questioning-Recall versus Categorical-Class and Questioning-Recall versus Attitude had negative correlations. Therefore, Questioning-Recall must be mutually exclusive to Categorical-Class and Attitude. Category Questioning-Recall must apply to formal and the categories Categorical-Class and Attitude must apply to concrete students.

Example 2: Categorical-Class versus Principles-Application had a negative correlation for concrete students.

Categorical-Class versus Principles had a negative correlation for concrete students.

Categorical-Class versus Application had a positive correlation for concrete students.

Example 1 showed Categorical-Class applied to concrete students. This example shows Categorical-Class must be encouraged through Application not through Principles. If Application applies to concrete students, then Principles must apply to post-concrete students.

Example 3: Categorical-Class versus Principles positive correlation for transitional students, but Categorical-Class versus Application had a zero correlation for transitional students. Compared to the example 2, Application no longer applies to transitional students and the relationship with Categorical-Class has been taken over by Principles.

Example 2 concluded Application applied to concrete students. This example shows Application does not and Categorical-Class and Principles do apply to transitional students. Apparently, the sequence developed is: A----CC----P, from concrete to transitional students.

By taking into account the results of Spearman rho and Kendall's W, as well as the inter-instrument correlation data, the following model was developed:

Part-Whole--> Whole-> Cat-Functional--> Cat-Class----> P-->
and Application R--->QR
------------------------> Enjoyment---------------> Attitude-->

PRE-CONCRETE (?) CONCRETE TRANSITIONAL FORMAL

The model was tested by dividing the students up into S.O.L.O. levels and further sub-dividing each S.O.L.O. level into successful and unsuccessful. Then, each S.O.L.O. level was categorized according to the instrument results. Thus, numbers of students in each instrument category at each S.O.L.O. level were determined. Next, the student numbers were tested, using F ratios (Dayton, 1970), for significant deviation from the number that would be expected at each sub-division . Table 6.13 below is presented as an illustration of how the F ratio was calculated.

Table 6.13 Calculating F-ratios for significance testing to determine if certain instrument categories are significantly different.

Instrument category	S.O.L.O. Multiple number of students in the categories unsuccessful	successful	significance
Principles	8	13	
Questioning	7	15	
Application	1	15	1 significantly low $p<0.03$
Questioning-Recall	7	11	11 below 1 std dev not significant
Principles-Application	10	13	10 significantly high $p<0.02$

After many calculations to determine the level of significance, the model was accepted largely intact with the exceptions of Enjoyment and Attitude. The position of these two instrument categories in the model could not be determined with any degree of significance. Thus, the model, at the end of the non-parametric analysis, appeared as follows:

Part----> Whole-->	Cat-Functional-> Cat-Class--------->	P------> R---> QR	
Whole	and Application		
	CONCRETE	TRANSITIONAL	FORMAL
S.O.L.O.	S.O.L.O.	S.O.L.O.	S.O.L.O.
Singular	Multiple	Relational	General Abstract

The parametric statistical analysis further refined this model. The justification for the refinements were presented in the section: Hypothesis 2(b). The final model was presented in table 6.11.

6.3 Concluding remarks

In the research done by Bates (1978, 58) the variances of the achievement scores in the inquiry method were found to be greater than those for the traditional teaching methods. Bates concluded that this greater variance was due to the important role the laboratory played in the inquiry method. In this study, Bates' results were not confirmed as the data in table 6.14 show.

Ramsey and Howe (1969, 64) and Bybee (1970, 160 and cited in Bates, 1978, 62) found in their studies that attitudes were the only significant differences between the laboratory and traditional methods with the laboratory method producing better attitudes. Because the laboratory played an important role in the inquiry method (Bates, 1978, 58), it could be argued that Ramsey, Howe and Bybee's observations should also hold for the inquiry method. Table 6.15 of class averages, as measured in this study, showed no significant differences in Attitude, as calculated by F ratios. The results in this study agreed with Yager, Engen and Snider (1969, 85) who also found no significant differences in attitude.

Table 6.14 Class variances of end of quarter achievement scores and selected instrument categories' scores for all treatment and control groups.

Class	Variances				
	achievement	Attitude	Enjoyment	Questioning -Recall	Principles -Application
Chemistry in the Market place	13.42	7.05	5.50	9.65	9.84
Advanced Science	19.88	7.12	5.63	11.77	6.85
Bubbles and Bangs	18.55	7.01	8.73	8.37	8.44
Experimental	18.31	5.86	8.75	11.32	5.64
Green Machine	25.83	6.15	4.64	12.03	6.54
Control	19.82	5.80	14.05	13.01	9.75

Note: to determine if the control group was significantly different in variance on these particular categories, F scores revealed only Enjoyment was significantly different in variance to a level of $p<0.02$.

Table 6.15 Class averages of achievement and selected instrument categories' scores for all treatment and control groups.

Class	Averages				
	achievement	Attitude	Enjoyment	Questioning -Recall	Principles -Application
Chemistry in the Market place	58.00	35.19	34.14	2.25	2.15
Advanced Science	47.96	39.88	39.24	-0.89	-4.06
Bubbles and Bangs	49.79	39.16	33.00	-2.23	1.69
Experimental	58.31	38.71	32.33	-2.96	-1.86
Green Machine	28.00	35.91	36.13	-5.82	0.89
Control	53.98	41.10	29.14	-1.59	0.32

The table of averages showed this study did not find a significant difference in achievement scores. This result agreed with the results from Zingaro and

Collette (Bates, 1978, 62), Bybee (1970, 160) and Yager, Engen and Snider (1969, 85), they also found no significant differences in knowledge acquisition. However, in a matched comparison between experimental and control groups, the experimental group outperformed the control group and with a lower variance, although the latter not to a significant level.

With regard to knowledge acquisition, Mulopo and Fowler (1978, 218) and Renner (1976, 222) concluded that formal reasoners benefitted most from the discovery and inquiry methods respectively. In this study, formal reasoners equated to S.O.L.O. Relational, Relational-General Abstract and General Abstract students. Based on Questioning-Recall, Principles-Application, Whole, Categorical-Functional, Categorical-Class, Attitude and Enjoyment scores and taking the five inquiry classes as a whole, S.O.L.O. Multiple students were 6.9% more likely to be successful than S.O.L.O. Relational students as calculated by the W.N.P.D. method. However, on a control versus experimental group basis, this difference was not significant for the experimental group at S.O.L.O. Multiple level. The experimental S.O.L.O. Multiple group's average was greater on criteria Pr-1, C-1, I-3 and K, but less on criteria Pr-2 and Pr-3. For S.O.L.O. Relational and General Abstract, the experimental group outscored the control group on all criteria to a significance level of $p<0.08$.

Further comparisons resulted in the W.N.P.D. values listed in the tables below.

Table 6.16 Selection of W.N.P.D. data showing the percentage likelihood of S.O.L.O. Multiple and Relational successful and unsuccessful students scoring above or below the average end of quarter achievement score.

Student category	S.O.L.O.	W.N.P.D. statement
successful	Multiple	3.3% more likely to score > av. than < av.
successful	Relational	-6.4% more likely to score > av. than < av.
unsuccessful	Multiple	-2.4% more likely to score > av. than < av.
unsuccessful	Relational	-10.0% more likely to score >av. than < av.

Legend: av.=average end of quarter achievement score
>=above
<=below

Table 6.17 General statements, as determined by W.N.P.D. calculations, relating to the claim that the inquiry method only favours formal reasoning students .

	Summary of W.N.P.D. statements	
Student category	S.O.L.O.	W.N.P.D. statement
all students	Relational	always more likely to score < av. than > av.
successful	Multiple	always more likely to score > av. than < av.
unsuccessful	Multiple	always more likely to score < av. than > av.

Legend: av.=average end of quarter achievement score
>=above
<=below

Therefore, although the inquiry method did not necessarily only favour formal reasoners, the results of this study agreed largely with Mulopo, Fowler and Renner: S.O.L.O. Relational and General Abstract students overall benefitted more from the inquiry, concrete based instruction than S.O.L.O. Multiple students. Lawson (1985, 604) found that concrete based instruction is beneficial to all students and the results of this study did not contradict his findings.

With regard to the percentage of students in each S.O.L.O. level McKinnon (1970, 72) determined that at College level 22% of students are formal, 27% are transitional and 51% operate at concrete level. In this study, for all the five inquiry classes combined, 6% of students are classified at formal level, 40% are transitional and 54% are concrete. A comparison of these data suggests that 13% of transitional students complete the transition to formal and 3% of concrete students develop in transitional students in a period of a few years. Although it must be acknowledged that the data do not come from the same student body, it is a matter of concern that so many students are classified at a concrete level (Collis and Briggs, 1989, 19).

Based on achievement scores, the inquiry method produced equal to or better results than the conventional teaching method. The inquiry method was especially successful for transitional and formal students. The S.P.C.S. instrument also showed that the difference between successful and unsuccessful students was less for high ability (transitional and formal) students than for low ability (concrete) students.

Mulopo and Fowler (1987, 218) and Renner (1976,222) asserted that formal reasoners would have more success in the laboratory and discovery methods. This research confirmed that for the inquiry method this was also the case. Egelston (1973, 476), Zingaro and Collette (cited in Bates, 1978, 62), Ramsey and Howe (1969, 64), Bybee (1970, 160) all reported no differences between laboratory, discovery and conventional teaching methods in knowledge acquisition. This research indicated that this was also the case for the inquiry method.

Ramsey and Howe (1969, 64), Sorenson (1966 and cited in Bates, 1978, 59), Bybee (1970, 160) and Mulopo and Fowler (1987, 218) all found that attitude was much better in laboratory and discovery methods. The S.P.C.S. instrument results indicated that successful students differed from unsuccessful students on methodology, planning ability in experimenting and ability to carry out an experiment. These factors were concerned with process. This backed up Nielsen's (1986, 1) and Janners' views (cited in Bates, 1978, 32) that science process was best taught in a laboratory environment, although this was doubted by Bates (1978, 55) and Ramsey and Howe (1969, 63) who were not convinced. Therefore, the inquiry method seemed to promote understanding science and, perhaps, openmindedness. These two factors were promoted in the laboratory method as reported by Sorenson (1966) and it seems justified to extend this to the inquiry method in light of the S.P.C.S. instrument's results. The hierarchy of instrument categories model for inquiry students showing the importance of particular categories for concrete, transitional and formal students indicated that a positive scientific attitude was desirable for formal and, perhaps, transitional students. This also supported the assertion that the inquiry method promoted understanding of science.

Lawson (1985, 604) asserted that concrete based instruction was beneficial to all students. This research indicates this was only partly true. Concrete and transitional students, most of the high school students belong to these two groups, did need science courses that emphasized Utility (concrete use and application) to increase the likelihood of success. Classification skills, which were also emphasized at transitional level, could also begin with concrete based instruction. However, formal students operate mainly at Principles and Recall levels and placed little emphasis on concrete based instruction.

Comparisons among the inquiry classes and between matched inquiry and control classes led to the development of a model showing at which of the three stages of development, viz. concrete, transitional and formal, particular skills or categories were emphasized. The model was supported by extensive statistical analysis and by the regression equations. The model accurately predicted the form of the regression equations for transitional students. The nine co-variable regression equation also proved accurate.

It must, however, be emphasized that the regression equations and the model should be seen as constraints not determinants for success in the inquiry method of teaching science.

6.3.1 Implication for science teachers

Three simple implications can be stated in the light of this research:
(a) for high ability students in particular the inquiry method of teaching science can be used without fear of disadvantaging these students;
(b) curriculum content does not need to be changed to allow the inquiry method of teaching science to be used;
(c) all students, but high ability students to a limited degree only, need help

with three factors which unsuccessful inquiry students identified as their major problem areas:

(1) experimental methodology, that is breaking down the experiment into a sequence of steps, what equipment to use, how much materials and equipment to use, etcetera;

(2) planning the experiment, that is what control to use, identification of variables, formulation of an aim and a hypothesis, methods of collecting results;

(3) guiding students with the execution of the experiment.

The first factor, listed under (c), can be largely overcome by familiarization with the phenomenon being studied and the equipment used. Through experience students will be able to judge, for instance, how quickly to heat glassware and how much much liquid is needed in a beaker in order to measure the liquid's boiling point. With regard to the second factor, planning experiments, students can be helped by giving them variations of the same phenomenon. In this way students will be better able to identify relevant and irrelevant variables based on what they observed in previous experiments. With experience comes confidence and the latter quality students need in order to bring their experiments to a conclusion. When students are aware of the relevant and irrelevant variables they are less likely to bog down in trivial problems that arise during the running of an experiment.

It is important that the teacher identifies the students' cognitive stage of development as early as possible. This identification does not really need to be done through the administration of tests because often teachers can be guided by their experience in classifying their students' stage of cognitive development as either concrete, transitional or formal.

The hierarchy of instrument categories model (refer table 6.14) can then be applied when planning a unit of work. To show explicitly the application of this model to the inquiry method of teaching science, in the following example the same science topic is used for a unit of work planned for students at the concrete stage of cognitive development and for students at the formal stage of cognitive development.

In this example, the science topic is pressure. Students at the concrete stage of cognitive development would benefit most if pressure were presented, or investigated, as one phenomenon and not as a relationship between force and area. Therefore, it would not be appropriate to do any experiments where forces and areas are measured and the pressure calculated. Instead, pressure applications would be more beneficial. Thus, U-tube experiments, balloon experiments, pumping arrangements together with either U-tubes or pressure gauges, tyre experiments, wind pressure investigations are all satisfactory as introductions.

Following these introductions, experiments concerned with categorizing applications by class are suggested by the model as appropriate. Categories could be classed by medium (for instance, fluid type: gas, liquid, plastic), by type of pump (for instance, reciprocating, rotary), applications of vacuum systems and applications of positive pressure systems. The next stage would be an appropriate starting point for formal students: the principle of pressure. At this stage the three variables: pressure, force and area, are investigated. Simple piston arrangements, pressure measurements with either U-tubes or pressure gauges, or both, at the bottom of fluid columns of different diameters, calculations of weight of a given amount of fluid and pressure exerted at the bottom of containers of different sizes, relationship between pressure and height in a column of fluid, mechanical advantage experiments that use fluids as working medium are all acceptable in the sense that their working principles are all founded on pressure, force and area.

Recall of facts, laws, phenomena, applications, etcetera, follow when the principle of pressure has been grasped. For instance, pressure, volume and temperature relationships for ideal fluids (usually air is acceptable) can be investigated. Investigations into hydraulic systems, the particle model of solids and fluids to explain pressure, venturis, whistles, relationships between work done, pressure, force and area are all appropriate for formal students. The culmination of a unit of work on pressure for formal students could be free inquiry, into facets of application or pressure phenomena, that are student initiated. Appropriate investigations could be made into: weather effects of pressure, Newtonian versus non-Newtonian fluids, labour saving devices (that is mechanical advantage systems), wind tunnel experiments, friction experiments, free fall and parachute experiments, nozzles and venturi applications, etcetera.

This example of a unit of work on pressure has been included to show the application of the hierarchy of instrument categories model developed as part of this thesis. Although this example would serve well as a list of content planned to be covered in the unit, to teach the unit using the inquiry method of teaching science, each activity, lesson, experiment and investigation would need to take into account the principle issues that are fundamental to the inquiry method. These are: process, no artificial separation of content and method, encourage higher level thinking, systematic experimentation, realistic conducting of inquiry into science, including both stable and fluid inquiry and ranging from closed-ended to open-ended inquiry, introduce discrepant events (events that appear to contradict predictions), encourage hunch generation and familiarizing students with a feeling of doubt (refer section *4.6.2 Aims page 90*). However, an emphasis on certain instrument categories must be taken into account and therefore modify these issues. The next few paragraphs further illustrate the use of the model, but in a less elaborate way.

Instrument categories have an effect on achievement scores and they are also unlikely to change in the time span of one term/ quarter. Further, they do not change as the result of the teaching method, at least not in the short term. The instrument categories are not related to cognitive stage of development, that is success is not determined by ability, as shown by the Chi-square test results (refer page 134). Success in the inquiry method is not related to cognitive stage of development. This means that some students who fail science when taught by the traditional method may do better when taught by the inquiry method. However, some students appeared unsuited to the inquiry method as it was taught in this study. This means that the inquiry method of teaching science, as suggested in the literature, must be amended to take into account the model developed in this study. The paragraphs below illustrate these amendments by highlighting the emphasis that must be put on certain instrument categories when using the inquiry method of teaching science.

The S.O.L.O. instrument results revealed that the bulk of grade nine students (S.O.L.O. Multiple) are incapable of generalizing relationships outside their immediate field of experience based on a series of data. For instance, if students were given three aqueous solutions of methylated spirits (metho) of equal volume but each with a different amount of water, they should be able to conclude that the more metho is diluted with water the higher its boiling point, but little more. About a quarter of grade nine and more than half of grade ten students (S.O.L.O. Relational) should be able to generalize within their field of experience. These students should be able to work out the boiling point of pure metho, but formulating the boiling point of any mixture (by taking into account part volumes and boiling points) would be beyond them. Using the model, most grade nine students would benefit from an emphasis on application and function of the boiling point of metho/water mixtures (burners, lantarns). Some grade nine and about half grade ten students could focus on, for instance, the implications of separating mixtures based on boiling points. The model suggests emphasizing both Categorical-Class and Categorical-Functional, for instance classing mixtures as fully, partially or non-separable as part of (industrial) distilling experiments. Late in grade ten general boiling point formulation could be attempted by students progressing to S.O.L.O. General Abstract.

This investigation into the relationship between the student characteristics and the inquiry method of teaching science has not only answered all hypotheses (refer sections 4.1.1 and 4.1.3) but it has also demonstrated that the inquiry method is a viable teaching method. When the conclusions of this research are used to modify the inquiry method (as used in this study) the statistical analyses have shown more students can successfully participate in the inquiry method. Even without modifications students studying by the inquiry method attained higher achievement scores on many of the Tasmanian Certificate of Education criteria than their traditionally taught peers and, overall, performed at least as well as those taught traditionally. This means that it is justified to argue that the inquiry method should be re-included in the arsenal of teaching methods available to science teachers, not at the exclusion of other methods but as another means of encouraging more students to be successful in science.

BIBLIOGRAPHY

Abraham, M.R. and Renner, J.W., 1986, 'The sequences of learning cycle activities in high school chemistry', Journal of Research in Science Teaching, 23, (2), 121-143.

Asher, J.W., 1976, Educational Research and Evaluation Methods, Little, Brown and Company, Boston.

Ausubel, P., 1968, 'An evaluation of the conceptual schemes approach to science curriculum development', Romery, W.D., Inquiry Techniques for Teaching Science, Prentice-Hall Inc., New Jersey.

Babikian, Y., 1971, 'An empirical investigation to determine the relative effectiveness of discovery, laboratory and expository methods of teaching science concepts', Journal of Research in Science Teaching, 8, (3), 201-209.

Bassler, O.C. and Kolb, J.R., 1971, Learning to Teach Secondary School Mathematics, Intext educational publishers, New York.

Bates, G.C., 1978, 'The role of the laboratory in secondary school science programs', Rowe, M.B. (ed.), What Research Says to the Science Teacher, National Science Teachers Association, Washington D.C.

Behr, M.J. and Eastman, P.M., 1978, 'Development and validation of two cognitive preference scales', Journal of Experimental Education, 46, (3), 28-34.

Biggs, J., 1989, 'Towards a model of school-based curriculum development and assessment using the S.O.L.O. taxonomy', Australian Journal of Education, 33, (2), 151-163.

Biggs, J. and Collis, K., 1989, 'Towards a model of school-based curriculum development and assessment using the S.O.L.O. taxonomy', Australian Journal of Education, 33, (2), 151-163.

Bruning, J.L. and Kintz, B.L., 1968, Computational Handbook of Statistics, Scott, Foresman and Company, Illinois.

Bruvold, W.H., Parlette, N., Bramson, R.M. and Bramson, S.J., 1983, 'An investigation of the item characteristics, reliability and validity of the inquiry mode questionaire',Educational and Psychological Measurement, 43, (2), 483-493.

Buckingham, J.P., 1975, 'Problems in logic in the classroom', The Science Teacher, 42, (1), 51-54.

Burns, R.W. and Ellis, B.M., 1970, 'What is discovery learning?', The American Biology Teacher, 54, (2), 105-107.

Bybee, R.W., 1970, 'The effectiveness of an individualized approach to a general education earth science laboratory', The American Biology Teacher, 54, (2), 157-161.

Cantu, L.L. and Herron, J.D., 1978, 'Concrete and formal piagetian stages and science concept attainment', Journal of Research in Science Teaching, 15, (2), 135-143.

Carter, G.E., 1982, 'Assessing students' interests in chemistry', British Journal of Educational Psychology, 52, (3), 378-380.

Case, R. and Fry, C., 1973, 'Evaluation of an attempt to teach scientific inquiry and criticism in a working class high school', Journal of Research in Science Teaching, 10, (2), 135-142.

Charles, C.M., 1976, The Instructional Endeavour, 2nd ed., The C.V. Mosby Company, Saint Louis.

Cheong, S., 1971, An Analysis of Inquiry Performance of High School

Biology Students Including the Relationship of Inquiry Performance to Instructional Techniques and to Student Achievement and Academic Ability, Ph. D. Dissertation, University of Texas, Austin, U.S.A.

Collis, K.F., 1990, Personal communication.

Collis, K.F. and Biggs, J.B., 1989, 'A school-based approach to seting and evaluating science curriculum objectives: S.O.L.O. and school science', The Australian Science Teachers Journal, 35, (4), 15-25.

Collis, K.F. and Davey, H.A., 1984, The Development of a Set of S.O.L.O. Items for High School Science, University of Tasmania, Tasmania.

Collis, K.F. and Davey, H.A., 1986, 'A technique for evaluating skills in high school science', Journal of Research in Science Teaching, 23, (7), 651-663.

Costenson, K. and Lawson, A.E., 1986, 'Why isn't inquiry used in more classrooms?', The American Biology Teacher, 48, (3), 150-158.

Coulter, J.C., 1966, 'The effectiveness of inductive laboratory, inductive demonstration and deductive laboratory in biology', Journal of Research in Science Teaching, 4, 185-186.

Creek, R.J. and Vollmer, M.L., 1991, The Inquiring School Model at Falk Laboratory School, Technical Report, University of Pittsburgh, Pennsylvania, U.S.A.

Davson-Galle, P., 1989, Personal communication.

Dayton, C.M., 1970, The Design of Educational Experiments, McGraw-Hill, New York.

Dishner, E.K. and Boothby, P.R., 1986, 'A bold "old" step: return to laboratory schools', New Directions for Teaching and Learning, (27), 49-57.

Driver, R., 1983, The Pupil as Scientists?, Open University Press, England.

Duschl, R.A., 1986, 'Textbooks and teaching of fluid inquiry', School Science and Mathematics, 86, (1), 27-32.

Egelston, J., 1973, 'Inductive versus traditional methods of teaching high school biology laboratory experiments', The American Biology Teacher, 57, (4), 467-477.

Fensham, P.J., 1981, 'Heads, hearts and hands - future alternatives for science education', The Australian Teachers Journal, 27, (1), 53-60.

Fish, A.S. and Goldmark, B., 1968, 'Inquiry method : three interpretations', Romey, W.D., Inquiry Techniques for Teaching Science, Prentice-Hall Inc., New Jersey.

Fisher, D., 1985 and 1986, Personal communication.

Foshay, A.W., 1991, 'Functions and purposes for campus laboratory schools', National Association of Laboratory Schools Journal, 15, (3), 3-16.

Friedler, Y. and Tamir, P., 1986, 'Teaching basic concepts of scientific research to high school students', Journal of Biological Education, 20, (4), 263-269.

Gallagher, J.J., 1987, 'A summary of research in science education - 1985', Science Education, 71, (3), 271-457.

Gardner, P.L. and Tamir, P., 1989, 'Interest in biology, part 1: a multidimensional construct', Journal of Research in Science Teaching, 26, (5), 409-423.

Glass, G.V. and Stanley, J.C., 1970, Statistical Methods in Education and Psychology, Prentice-Hall Inc., New Jersey.

Good, T.L. and Brophy, J.E., 1978, Looking in Classrooms, 2nd ed., Harper and Row, New York.

Good, R., Kromhout, R.A. and Mellon, E.K., 1979, 'Piaget's work and chemical education', Journal of Chemical Education, 56, (7), 426-430.

Guilford, J.P., 1965, Fundamental Statistics in Psychology and Education, McGraw-Hill, New York.

Guilford, J.P. and Fruchter, B., 1973, Fundamental Statistics in Psychology and Education, 5th ed., McGraw-Hill, New York.

Hall, D.A. and McCurdy, D.W., 1990, 'A comparison of a biological sciences curriculum study (B.S.C.S.) laboratory and a traditional laboratory on student achievement at two private liberal arts colleges', Journal of Research in Science Teaching, 27, (7), 625-636.

Harpole, S.H., 1987, 'The relationship of gender and learning styles to achievement and laboratory skills in secondary school chemistry students', Paper presented at the Annual Meeting of the Mid-South Educational Research

Association (Mobile, Al., November).

Harris, C.W. (ed.), 1960, Encyclopedia of Educational Research, 3rd ed., The Macmillan Company, New York.

Harty, H., Hamrick, L. and Samuel, K.V., 1985, 'Relationships between middle school students science concept structure interrelatedness competence and selected cognitive and affective tendencies', Journal of Research in Science Teaching, 22, (2), 179-191.

Hayslett, H.T., 1981, Statistics Made Simple, William Heinemann, London.

Heath, R.W., 1964, 'Curriculum, cognition, and educational measurement', Educational and Psychological Measurement, 24, (2), 239-253.

Herman, G.D. and Hincksman, N.G., 1978, 'Inductive versus deductive approaches in teaching a lesson in chemistry', Journal of Research in Science Teaching, 15, (1), 37-42.

Herron, J. D.,1975, 'Piaget for chemists : explaining what "good" students cannot understand', Journal of Chemical Education, 52, (2), 146-150.

Herron, J.D., 1977, 'Using chemistry problems to provoke self-regulation', Journal of Chemical Education, 54, (1), 41-43.

Hofstein, A. and Lunetta, V.N., 1982, 'The role of the laboratory in science teaching : neglected aspects of research', Review of Educational Research, 52, (2), 201-217.

Hounshell, P.B., 1985, The Science Laboratory; Research Report Number Six, North Carolina University, Scool of Education, Chapel Hill, U.S.A.

Humphreys, D.W., 1978, 'Teaching by inquiry', The American Biology Teacher, 40, (7), 435-437.

Igelsrud, D. and Leonard, W.H., 1988, 'What research says about biology laboratory instruction', The American Biology Teacher, 50, (5), 303-306.

Ivins, J.E., 1985, A Comparison of the Effects of Two Instructional Sequences Involving Science Laboratory Activities, Ph. D. Dissertation, University of Cincinnati, U.S.A.

Janners, M.Y., 1988, 'Inquiry, investigation and communication in the student-directed laboratory, Journal of College Science Teaching, 18, (1), 32-35.

Jungwirth, E., 1980, 'Alternative interpretations of findings in cognitive preference research in science education', Science Education, 64, (1), 85-94.

Khalili, K.Y., 1987, 'A crosscultural validation of a test of science related attitudes', Journal of Research in Science Teaching, 24, (2), 127-136.

Kolb, D.A., Rubin, I.M. and McIntyre, J.M., 1974, Organizational Psychology : A Book of Readings, 2nd ed., Prentice-Hall Inc., New Jersey.

Kraft, R.G., 1985, 'Group-inquiry turns passive students active', College Teaching, 33, (4), 149-154.

Lawson, A.E., 1978, 'The development and validation of a classroom test of formal reasoning', Journal of Research in Science Teaching, 15, (1), 11-24.

Lawson, A.E., 1985, 'A review of research on formal reasoning and science teaching', Journal of Research in Science Teaching, 22, (7), 569-618.

Lawson, A.E., 1988, 'A better way to teach biology', American Biology Teacher, 50, (5), 266-274.

Lawson, A.E., Blake, A.J.D. and Nordland, F.H., 1975, 'Training effects and generalization of the ability to control variables in high school biology students', Science Education, 59, (3), 387-396.

Lawson, A.E. and Blake, A.J.D., 1976, 'Concrete and formal thinking abilities in high school biology students as measured by three separate instruments', Journal of Research in Science Teaching, 13, (3), 227-235.

Lawson, A.E., Nordland, F.H. and DeVito, A., 1975, 'Relationships of formal reasoning to achievement, aptitudes and attitudes in preservice teachers', Journal of Research in Science Teaching, 12, (4), 423-431.

Lawson, A.E. and Renner, J.W., 1975, 'Relationships of science subject matter and development levels of learners', Journal of Research in Science Teaching, 12, (4), 347-358.

Lawson, A.E. and Smitgen, D.A., 1982, 'Teaching formal reasoning in a college biology course for preservice teachers', Journal of Research in Science Teaching, 19, (3), 233-248.

Lawson, A.E. and Wallman, W.T., 1976, 'Encouraging the transition from concrete to formal cognitive functioning : an experiment', Journal of Research in Science Teaching, 13, (5), 413-430.

Leonard, W.H., 1980, 'Using the extended discretion approach to biology laboratory investigation', The American Biology Teacher, 42, (7), 338-341.

Leonard, W.H., 1984, 'An experimental test of an extended discretion laboratory approach for university general biology', Paper presented at the Annual Meeting of the National Association for Research in Science Teaching (57th, New Orleans, April 29, 1984).

Leonard, W.H., Cavana, G.R. and Lowery, L.F., 1981, 'An experimental test of an extended discretion approach for high school biology laboratory investigations', Journal of Research in Science Teaching, 18, (6), 497-504.

Lumpe, A.T. and Oliver, J.S., 1991, 'Dimensions of hands-on science', The American Biology Teacher, 53, (6), 345-348.

Lunetta, V.N. and Tamir, P., 1981, 'An analysis of laboratory activities: project physics and P.S.S.C.', School, Science and Mathematics, 81, (8), 635-642.

McKinnon, J.W., 1970, The Influence of a College Inquiry-Centered Course in Science on Student Entry into the Formal Operational Stage, Ph. D. Dissertation, The University of Oklahoma, U.S.A.

McNaught, C., 1982, 'Relationship between cognitive preferences and achievement in science', Journal of Research in Science Teaching, 19, (2), 177-186.

Mills, V.M., 1981, 'The investigative laboratory in introductory biology courses : a practical approach', The American Biology Teacher, 43, (7), 364-367.

Mulopo, M.M. and Fowler, H.S., 1987, 'Effects of traditional and discovery instructional approaches on learning outcomes for learners of different intellectual development : a study of chemistry students in Zambia', Journal of Research in Science Teaching, 24, (3), 217-227.

Nash, B.J. and Hargreaves, G.J., 1990, Chemistry Activities, The Macmillan company of Australia, South Melbourne.

Nielsen, R.A., 1986, 'Laboratory schools: blue print for success', Paper presented at the Annual Convention of the National Association of Laboratory Schools (Chicago, Il., February 23-27), U.S.A.

Novak, J.A. and Voss, B.E., 1981, 'An investigation of relationships between cognitive preference orientation and Jungian (M.B.T.I.) personality types of eigth-grade science students', Paper presented at the Annual Meeting of the National Association for Research in Science Teaching (54th, Grossinger's in the Catskills, Ellenville, New York, April 5-8).

Ogunyemi, E.L., 1973, 'Cognitive styles and student science achievement in Nigeria', Journal of Experimental Education, 42, (1), 59-63.

Okebukola, P.A., 1985, 'Science laboratory behaviour strategies of students relative to performance in and attitude to laboratory work', Journal of Research in Science Teaching, 22, (3), 221-232.

Okebukola, P.A., 1986, 'An investigation of some related factors affecting students' attitudes toward laboratory chemistry', Journal of Chemical Education, 63, (6), 531-532.

Okebukola, P.A., 1986, 'The influence of preferred learning styles on cooperative learning in science', Science Education, 70, (5), 509-517.

Okebukola, P.A., 1987, 'Students' performance in practical chemistry: a study of some related factors', Journal of Research in Science Teaching, 24, (2), 119-126.

Okebukola, P.A. and Jegede, O.J., 1988, 'Cognitive preference and learning mode as determinants of meaningful learning thruogh concept mapping', Science Education, 72, (4), 489-500.

Page, J.A., 1983, 'Laboratory schools: updated or outdated?', Education, 103, (4), 372-374.

Pallett, R., 1983, Skills and S.O.L.O in Science, Tasmanian Education Department, Hobart.

Pella, M.O. and Sherman, J., 1969, 'A comparison of two methods of utilizing laboratory activities in teaching the course I.P.S.', School, Science and Mathematics, 69, (4), 303-314.

Peterson, K., 1978, 'Learning about subject areas: science inquiry', Journal of Research in Science Teaching, 15, (3), 227-232.

Pickering, M., 1985, 'Laboratory is a puzzle, not an illustration', Journal of Chemical Education, 62, (10), 874-875.

Pickering, M. and Crabtree, R.H., 'How students cope with a procedureless laboratory exercise', Journal of Chemical Education, 56, (7), 487-488.

Pilliner, A., 1975, Educational Studies : A Third Level Course, Methods of Educational Enquiry, Block 5 : Experiments in Educational Research, The Open University Press, Great Britain.

Pugliese, A.C., 1973, 'The meaning of inquiry, discovery, and investigative approachesto science teaching', Science Teacher, 40 (1), 24-27.

Purser, R.K. and Renner, J.W., 1983, 'Results of two tenth-grade biology teaching procedures', Science Education, 67, (1), 85-98.

Ramsey, G.A. and Howe, R.W., 1969, 'An analysis of research on instructional procedures in secondary school science : part I outcomes of instruction', The Science Teacher, 36, (3), 62-71.

Ramsey, G.A. and Howe, R.W., 1969, 'An analysis of research on instructional procedures in secondary school science : part II. instructional procedures', The Science Teacher, 36, (4), 72-81.

Renner, J.W., 1976, 'Significant physics content and intellectual development - cognitive development as a result of interacting with physics content', American Journal of Physics, 44, (3), 218-222.

Renner, J.W., 1979, 'The relationships between intellectual development and written response to science questions', Journal of Research in Science Teaching, 16, (4), 279-299.

Renner, J.W. and Cate, J.M., 1985, 'Measured formal thought and that required to understand formal concepts in secondary school biology', Paper presented at the Annual Meeting of the National Association for Research in Science Teching (58th, French Lick Springs, I.N., April 15-18).

Renner, J.W. and Paske, W.C., 1977, 'Comparing twp forms of instruction in college physics', American Journal of Physics, 45, (9), 851-859.

Richardson, V. and Renner, J.W., 1970, 'A study of the inquiry-discovery method of laboratory instruction', Journal of Chemical Education, 47, (1), 77-79.

Roadrangka, V. and Yeany, R.H., 1982, 'A study of the relationship among type and quality of implementation of science teaching strategy, student formal reasoning ability, and student engagement', Paper presented at the National Conference of the associationfor the Education of Teachers in Science (Chicago, Il., April 3).

Romey, W.D., 1968, Inquiry Techniques for Teaching Science, Prentice-Hall Inc., New Jersey.

Roscoe, J.T., 1975, Fundamental Research Statistics for the Behavioural Sciences, 2nd ed., Holt, Rinehart and Winston Inc., New York.

Rost, J., 1983, 'Evaluation studies: cognitive preferences as components of student interest', Studies in Educational Evaluation, 9, (3), 285-302.

Rotheram, K., 1984, 'Guided exploration using flowcharts', School Science Review, 65, (233), 655-669.

Rotheram, K., 1987, 'Guided investigations', School Science Review, 68, (245), 631-634.

Rowe, M.D. (ed.), 1982, Education in the 80's: Science, National Education Association, Washington, D.C., U.S.A.

Rubin, A. and Tamir, P., 1988, 'Meaningful learning in the school laboratory', The American Biology Teacher, 50, (5), 477-482.

Rutherford, F.J., 1964, 'The role of inquiry in science teaching', Journal of Research in Science Teaching, 2, 80-84.

Rutherford. F.J., 1968, 'The role of inquiry in science teaching', Romey, W.D., Inquiry Techniques for Teaching Science, Prentice-Hall Inc., New Jersey.

Sakmyser, D.D., 1974, 'Comparison of inductive and deductive programmed instruction on chemical equilibrium for high school chemistry students', Journal of Research in Science Teaching, 11, (1), 67-77.

Schneider, L.S. and Renner, J.W., 1980, 'Concrete and formal teaching', Journal of Research in Science Teaching, 17, (6), 503-517.

Schwab, J.J., 1966, 'The teaching of science as enquiry', Schwab, J.J. and Brandwein, P.F., The Teaching of Science, Harvard University Press, Massachusetts.

Scott, N.C., 1964, 'Science concept achievement and cognitive functions', Journal of Research in Science Teaching, 2, 7-16.

Scott, N.C., 1966, 'The strategy of inquiry and styles of categorization', Journal of Research in Science Teaching, 4, 143-153.

Scott, N.C., 1970, 'Strategy of inquiry and styles of categorization : a three-year exploratory study', Journal of Research in Science Teaching, 7, 95-102.

Scott, N.C., 1973, 'Cognitive style and inquiry strategy : a five-year study', Journal of Research in Science Teaching, 10, (4), 323-330.

Seymour, L.A., Padberg, L.F., Bingham, R.M. amd Koutnik, P.G., 1974, 'A successful inquiry methodology', The American Biology Teacher, 36, (6), 349-352.

Shemesh, M. and Lazarowitz, R., 1985, Formal Reasoning Skills of Secondary School Students as Related to gender, Age, School Type and Learning Ability, Wolf Foundation and Technion, Grant number 230-098, Israel.

Sherman, L.W., 1989, 'A comparative study of cooperative and competitive achievement in two secondary biology classrooms: the group investigation model versus an individually competitive goal structure', Journal of Research in Science Teaching, 26, (1), 55-64.

Shuaiba, M.J. and Ogunsola, M.F., 1983, 'Cognitive styles in students of chemistry in S.B.S.-A.B.U. Zaria, Nigeria', Research in Science and Technological Education, 1, (1), 101-107.

Shulman, L.S. and Tamir, P., 1973, 'Research on teaching in the natural sciences', Travers, R.M.W. (ed.), Second Handbook of Research on Teaching; a Project of the American Educational Research Association, Rand McNally and Company, Chicago.

Shymansky, J.A., 1972, A Comparative Laboratory Study of the Effects of Two Teaching Pattens on Certain Aspects of the Behavior of Students in fifth grade Science, Ph. D. Dissertation, Florida State University, U.S.A.

Shymansky, J.A. and Yore, L.D., 1980, 'A study of teaching strategies, student cognitive development, and cognitive style as they relate to student achievement in science', Journal of Research in Science Teaching, 17, (5), 369-382.

Shymansky, J.A. and Penick, J.E., 1981, 'Teacher behavior does make a difference in hands-on science classrooms', School Science and Mathematics, 81, (5), 412-422.

Smith, G.M., 1970, A Simplified Guide to Statistics, 4th ed., Holt, Rinehart and Winston Inc., New York.

Soloman, J., 1980, Teaching Children in the Laboratory, Biddles Ltd., Surrey, Great Britain.

Sorenson, L.L., 1966, Change in Critical Thinking Between Students in Laboratory-Centered and Lecture-Centered Patterns of Instruction in High School Biology, Dissertation Abstract International, 26, 6567A.

Sternberg, R.J., 1986, 'Toward a unified theory of human reasoning', Intelligence, 10, (4), 281-314.

Sugrue, B.M. and Thomas, R.A., 1989, 'Effects of discovery and expository instruction on recall and transfer of procedural knowledge: interactions with learner aptitude', Paper presented at the Annual Meeting of the American Educational Research Association (San Fransisco, C.A., March 27-31).

Sund, R.B. and Trowbridge, L.W., 1967, Teaching Science by Inquiry in the Secondary School, Charles E. Merrill Publishing Company, Colombus, Ohio.

Talton, E.L. and Simpson, R.D., 1987, 'Relationships of attitude toward classroom environment with attitude toward and achievement in science among tenth grade biology students', Journal of research in Science Teaching, 24, (6), 507-525.

Tamir, P., 1976, 'Factors which influence student achievement in high school biolgy', Journal of Research in Science Teaching, 13, (6), 539-545.

Tamir, P., 1977, 'Students' characteristics and instructional outcomes', Journal of Research in Science Teaching, 14, (5), 477-480.

Tamir, P., 1978, 'Are cognitive preferences just an expression of cognitive ability?', Journal of Experimental Education, 46, (2), 60-65.

Tamir, P., 1983, 'Cognitive preferences of Jewish and Arab high school students who studied an inquiry oriented biology curriculum for several years', Research in Science and Technological Education, 1, (1), 17-26.

Tamir, P., 1985, 'Content analysis focussing on inquiry', Journal of Curriculum Studies, 17, (1), 87-94.

Tamir, P., 1985, 'Meta-analysis of cognitive preferences and learning', Journal of Research in Science Teaching, 22, (1), 1-17.

Tamir, P., 1988, 'The relationship beetween cognitive preferences, student background and achievement in science', Journal of Research in Science Teaching, 25, (3), 201-216.

Tamir, P. and Amir, R., 1987, 'Interrelationships among laboratory process skills in biology', Journal of Research in Science Teaching, 24, (2), 137-144.

Tamir, P. and Cohen, S., 1980, 'Factors that correlate with cognitive preferences of medical school teachers', Journal of Educational Research, 74, (2), 69-74.

Tamir, P. and Glassman, F., 1971, 'A practical examination for B.S.C.S. students: a progress report', Journal of Research in Science Teaching, 8, (4), 307-315.

Tamir, P. and Jungwirth, E., 1984, 'Test scores and associations as measures of cognitive preferences', Studies in Educational Evaluation, 10, (2), 149-158.

Tamir, P. and Kempa, R.F., 1978, 'Cognitive preference styles across three science disciplines', Science Education, 62, (2), 143-152.

Tamir, P. and Lunetta, V.N., 1978, 'Cognitive preferences in biology of a group of talented high school students', Journal of Research in Science Teaching, 15, (1), 59-64.

Tamir, P. and Lunetta, V.N., 1981, 'Inquiry-related tasks in high school science laboratory handbooks', Science Education, 65, (5), 477-485.

Tamir, P., Penick, J.E. and Lunetta, V.N., 1982, 'Cognitive preferences and creativity : an exploratory study', Journal of Research in Science Teaching, 19, (2), 123-131.

Thorndike, R.M., 1988, 'Correlational procedures', Keeves, J.P. (ed.), Educational Research, Methodology and Measurement; an international handbook, Pergamon Press, Sydney.

Tinnesand, M. and Chan, A., 1987, 'Step 1 : throw out the instructions', Science Teacher, 54, (6), 43-45.

Van Den Berg, E., Lunetta, V.N. and Tamir, P., 1978, 'Cognitive preferences: a validation study', Studies in Educational Evaluation, 4, (2), 107-120.

Van Den Berg, E., Lunetta, V.N. and Tamir, P., 1982, 'The convergent validity of the cognitive preference construct', Journal of Research in Science Teaching, 19, (5), 417-424.

Welch, W.W., Klopfer, L.E., Aikenhead, G.S. and Robinson, J.T., 1981, 'The role of inquiry in science education : analysis and recommendations', Science

Education, 65, (1), 33-50.

Westbury, I. and Wilkof, N.J. (eds), 1978, Joseph J. Schwab : Science, Curriculum and Liberal Education, Selected Essays, The University of Chicago Press, Chicago.

Wilkinson, J.W. and Nash, B.J., 1989, World of Chemistry, The Macmillan company of Australia, South Melbourne.

Wilson, J.T., 1974, 'Processes of scientific inquiry : a model for teaching and learning science', Science Education, 58, (1), 127-133.

Wilson, M., 1989, 'A comparison of deterministic and probabilistic approaches to measuring learning structures', Australian Journal of Education, 33, 2, 127-140.

Wollman, W. and Lawson, A.E., 1977, 'Teaching procedure of controlled experimentation : a piagetian approach', Science Education, 61, (1), 57-70.

Wollman, W.T. and Lawson, A.E., 1978, 'The influence of instruction on proportional reasoning in seventh grade', Journal of Research in Science Teaching, 15, (3), 227-232.

Yager, R.E., Engen, H.B. and Snider, B.C.F., 1969, 'Effects of the laboratory and demonstration methods upon the outcomes of instruction in secondary biology', Journal of Research in Science Teaching, 6, (1), 76-86.

APPENDIX A

The Cognitive Style of Classification Behaviour (C.S.C.B.) instrument

Introduction to the teacher:

In this instrument students are presented with three familiar things. While all three are related in some way, students are asked to select two only which they think are related; they show this by ticking the two selected items.

In the space on the right, students are asked to write how the two selected items are related. The reason does not have to be factually correct, it is sufficient if the student thinks it is correct.

Hand out the forms to the students.

Read out aloud to the students the following three examples:

1. Book,magazine,newspaper / /	-both have separate articles; -both are serials; -an article in a newspaper may be expanded on in the magazine; -both are thrown out after reading; -they are both thin; -both are meant to entertain -they are both paperbacks.
2. Hammer, fork, knife / /	-both have narrow tips; -both are all metal; -one holds the food while the other cuts it;

	-both are used to eat with; -both are cutlery; -we use both everyday.
3. Computer,telephone,t.v. / /	-both have screens; -both are larger than the phone; -the t.v. can be used as a screen for the computer; -both display information; -both are electronic; -both are modern inventions.

To the student.

Each question presents three things.

Tick (/) two which you think are related, but once ticked do not change your mind (your first choice is best).

On the right, write down why you think the two are related. You may write as many reasons as you can think of. If you cannot think of a reason skip the question and do the next one.

There are no right or wrong answers!

1. COWBOY	NURSE	POLICEMAN	
2. CAR	UTE	TRUCK	
3. CUP	GLASS	MUG	
4. COW		DOG	KANGAROO
5. T.V.	RADIO	VIDEO	
6. LAMP	TORCH	FLUORESCENT TUBE	
7. LOUNGE CHAIR	STOOL	PARK BENCH	
8. AXE		HAMMER	SAW
9. BOXER	RUNNER	TENNIS	

		PLAYER	
10 PLANT	TREE	FLOWER	
11 RECORD	CASSETTE	COMPUTER DISK	
12 MATCHES	LIGHTER	WOODEN STICK	
13 JUMPER	T-SHIRT	SINGLET	
14 SHORTS	JEANS	TRACKSUIT PANTS	
15 CAT	TIGER	DOG	
16 HAT	BEANIE	HELMET	
17 SUGAR	FLOUR	COCONUT	
18 HONEY	JAM	VEGEMITE	
19 NECKLACE	RING	BRACELET	
20 GUMBOOT		SNEAKER	SLIPPER
21 COKE	MILK	TEA	
22 FERN	ROSE	DAISY	
23 CARPET	LINO	CURTAINS	
24 TOWEL	SHEET	TABLECLOTH	
25 ZIP	BUTTON	VELCRO	
26 APRON	GLOVES	HAT	
27 SPIDER	FLY	SILVERFISH	
28 TADPOLE	BEE	CATERPILLAR	
29 YOGHURT	CHEESE	ICECREAM	

30 DRAWING PIN	PAPER CLIP	HAIR SLIDE

Key-words for student classification

Below are presented the key-words that were used to classify student responses in the various categories.

Descriptive Part-whole attribute:

needs water, runny, slippery, sweet, cold/warm, sticky, in containers, made from ..., has a flame, runs on batteries, hit things, no arms, battery operated.

Descriptive Whole attribute:

well paid, both run, both crawl, both tame, found outside, help people, play music, make fire, use electricity, shape, grow up, cover up legs, size, natural, hunt, go around something, are electric, go on the floor, give light, wear them on your head.

Relational-Contextual or Functiona- Interdependence:

used together, need each other, go together, one needs the other, milk goes in tea, can be put together, need one to light the other, if you froze yoghurt it would be like icecream.

Categorical-Functional:

worn for hot weather, are for sport, eaten on toast, used on paper, hold things, carry things, light fires (note difference with: make fire), carry goods, can watch them, for hot drinks, listen to them, keep head warm, lay them over things, protect.

Categorical-Class:

work for government, jobs are to help (note difference with: help people), hunters, live on a farm, are inside, inventions, floor coverings, undergarments, mammals, both cats, dairy products, both change into adults, contact sports, refrigerated, seat more than one.

Categorical-Inferential or Categorized by Irrelevancy:

disgusting, are everywhere, break easily, worn more often, more interesting, don't like fighting.

APPENDIX B

Details of the Weighted Net Percentage Difference (W.N.P.D.) analysis

The test categories that were included in the W.N.P.D. table were those categories that were shown to have significant relationships with either the successful\unsuccessful classification or with the criterion score. The following table resulted.

Table B.1 Maximum student numbers possible in the W.N.P.D. cells.

criterion score	teacher obs.	SOLO level	Q-R x>x	Q-R x<x	P-A x>x	P-A x<x	W x>x	CF x>x	CC x>x	At x>x	E x>x
		Singular	0	0	0	0	0	0	0	0	0
		Multiple	18	18	18	18	17	17	17	18	18
	success	Relational	17	17	17	17	17	17	17	17	17
	-ful	General									
		Abstract	3	3	3	3	3	3	3	3	3
above the mean											
		Singular	0	0	0	0	0	0	0	0	0
		Multiple	4	4	4	4	5	5	5	5	5
	unsucc	Relational	7	7	7	7	6	6	6	6	6
	-essful	General									
		Abstract	0	0	0	0	0	0	0	0	0
		Singular	0	0	0	0	0	0	0	0	0
		Multiple	9	9	9	9	11	11	11	11	11
	success	Relational	4	4	4	4	4	4	4	2	2
	-ful	General									
		Abstract	2	2	2	2	2	2	2	1	1
below the mean											
		Singular	0	0	0	0	0	0	0	0	0
		Multiple	9	9	9	9	11	11	11	11	11
	unsucc	Relational	4	4	4	4	6	6	6	5	5
	-essful	General									
		Abstract	0	0	0	0	0	0	0	0	0

Legend: teacher obs.=classification based on teacher observation throughout the quarter.

criterion score=either higher than the average or below the average criterion score for all classes as a whole.

Q-R=Questioning-Recall

P-A=Principles-Application

W=Whole

CF=Categorical-Functional

CC=Categorical-Class

At=Attitude

E=Enjoyment

The numbers table B.1 refer to the number of students who ought to be in that cell according to the S.O.L.O. classification scheme. The number along any row are not necessarily the same because not all students completed all the tests. For instance, there were 18 students who scored above the criterion score mean and rated S for teacher observation and were classed S.O.L.O. M. However, of these 18 students only 17 completed the S.C.C.B. test, but all 18 completed the C.C.P.I. test. The W.N.P.D. calculations do not use actual numbers but use percentage of students who actually fitted in the cells. For instance, of the maximum of 18 students from the example above only eight, or 44%, could actually be classified in the first cell of the second row. The table below shows the same cells as the previous table, but instead of showing maximum student numbers, percentages of students, relative to the maximum possible, who actually fitted in the cells are shown. These percentages are used in the W.N.P.D. calculations.

Table B.2 Percentage of students in the W.N.P.D. cells.

criterion score	teacher obs.	SOLO level	Q-R x>x	Q-R x<x	P-A x>x	P-A x<x	W x>x	CF x>x	CC x>x	At x>x	E x>x
		Singular	0	0	0	0	0	0	0	0	0
		Multiple	44	50	50	50	47	65	65	61	72
	success	Relational	47	53	71	29	35	41	41	59	65
	-ful	General									
		Abstract	67	33	67	33	33	0	100	67	33
above the mean											
		Singular	0	0	0	0	0	0	0	0	0
		Multiple	25	75	75	25	20	80	60	40	0
	unsucc	Relational	43	43	43	43	50	50	67	17	33
	-essful	General									
		Abstract	0	0	0	0	0	0	0	0	0
		Singular	0	0	0	0	0	0	0	0	0
		Multiple	33	67	44	56	45	45	55	80	50
	success	Relational	75	25	75	25	75	25	50	100	50
	-ful	General									
		Abstract	0	100	0	100	50	50	100	0	100
below the mean											
		Singular	0	0	0	0	0	0	0	0	0
		Multiple	67	33	78	22	60	30	50	40	40
	unsucc	Relational	25	100	75	50	50	33	33	60	60
	-essful	General									
		Abstract	0	0	0	0	0	0	0	0	0

Examples of comparisons that were made using the W.N.P.D. table

A total of 55 comparisons were made; many of these consisted of parts a and b. An example of one comparison is: how many times is a Successful student more likely to score above the criterion score average than below it? This compares a block of the first 36 cells (above average S students) to another block of 36 cells: cells 73 through to 108 inclusive (below average S students). The actual comparison only pits equivalent cells against each other. Thus, cell 1 is compared to cell 73. Cell 1 is the above average S student rated S on the S.O.L.O. test and scored above average on the Q-R category. Cell 73 is the below average S student rated S on the S.O.L.O. test and scored above average on the Q-R category. Cells 1 to 9 are all zero, as are cells 73 to 81; therefore, these cells need not be considered. The actual W.N.P.D. calculation is performed as shown in table B.3.

Table B.3 The actual W.N.P.D. calculation for an example comparison.

% diff.	cell no's.	total number of students	% x no. of students
11	10 - 82	27	297
-17	11 - 83	27	-459
6	12 - 84	27	162
-6	13 - 85	27	-162
2	14 - 86	28	56
20	15 - 87	28	560
10	16 - 88	28	280
-19	17 - 89	28	-532
22	18 - 90	28	616
-28	19 - 91	21	-588
28	20 - 92	21	588
-4	21 - 93	21	-84
4	22 - 94	21	84
-40	23 - 95	21	-840
16	24 - 96	21	336
-9	25 - 97	21	-189
-41	26 - 98	19	-779
15	27 - 99	19	258
67	28 -100	7	469
-67	29 -101	7	-469
67	30 -102	7	469
-67	31 -103	7	-469
-17	32 -104	7	-119
-50	33 -105	7	-350
0	34 -106	7	0
67	35 -107	4	268
-67	36 -108	4	-268
		sum= 490	sum= -838

W.N.P.D.= -838/490= -1.71%; that is, S students are 1.7% more likely to score below the average than above it based on all the test categories.

A similar calculation for Unsuccessful students resulted in a W.N.P.D. value of -6.49%. Clearly, the calculations can be modified to include only specific test categories. If, for instance, the comparison had been formulated as: how many times is a Successful student more likely to score above the criterion score average than below it, based on above average scores in all test categories (exclude Q-R X<X and P-A X<X), the W.N.P.D. value would

have been 0.13%.

Rather than a block by block comparison, a row by row comparison can be made by stipulating a specific S.O.L.O. level. In total, well over a hundred comparisons were made to determine to what extent Successful students were different from Unsuccessful students and on which test categories. Also, comparisons were made to find out the extent to which, for instance, Q-R X>X (above average) students were different from Q-R X<X (below average) students.

Results of comparisons that were made using the W.N.P.D. table

The comparisons were always worded in the following standard sentence:

(student category) students are ....% more likely to score (above or below) the criterion average score than (above, below or student category), (any restrictions on instrument categories).

Due to the large number of comparisons, only those comparisons with percentages greater than 15% are listed here. It must be realized however that percentages of zero or close to zero can also hold valuable information, to show, for instance, that two categories are either equally important or equally unimportant

.

The original comparison number has been maintained to facilitate using any results of this research not listed in this Appendix in any future research.

33.For S.O.L.O. M level, U students are 33.16% more likely to score above the criterion average score than below it (Q-R X>X, P-A X<X, CF, At).
44.For S.O.L.O. R level, U students are 27.73% more likely to score below the criterion average score than above it (Q-R X>X, P-A X>X, CF, At).
46.For S.O.L.O. M level, S students are 15.26% more likely to score above the criterion average score than U students (all X>X categories).
47.For S.O.L.O. M level, S students are 16.11% more likely to score above the criterion average score than U students (Q-R X<X, P-A X<X, W, CF, CC, At, E).
48.For S.O.L.O. M level, S students are 22.31% more likely to score above the criterion average score than U students (Q-R X>X, P-A X<X, W, CF, CC, At, E).
53.For S.O.L.O. M level, S students are 16.76% more likely to score below the criterion average score than U students (Q-R X<X, P-A X<X, W, CF, CC, At, E).
56.For S.O.L.O. M level, S students are 30.38% more likely to score below the criterion average score than U students (Q-R X<X, P-A X<X, CF, At).
62.For S.O.L.O. R level, S students are 17.78% more likely to score above

the criterion average score than U students (Q-R X<X, P-A X>X, CF, At).

64.For S.O.L.O. R level, S students are 15.83% more likely to score below the criterion average score than U students (all X>X categories).

68.For S.O.L.O. R level, S students are 18.18% more likely to score below the criterion average score than U students (Q-R X>X, P-A X>X, CF, At).

78.S students are 23.54% less likely to score below the criterion mean score than U students (P-A X>X).

80.U students are 16.29% less likely to score above the criterion mean score than below it (P-A X>X).

82.S students are 15.85% more likely to score below the criterion mean score than U students (P-A X<X).

88.U students are 35.33% more likely to score above the criterion mean score than below it (CF X>X).

89.S students are 31.50% more likely to score above the criterion mean score than U students (At X>X).

90.S students are 40.00% more likely to score below the criterion mean score than U students (At X>X).

91.S students are 20.45% less likely to score above the criterion mean score than below it (At X>X).

92.U students are 18.19% less likely to score above the criterion mean score than below it (At X>X).

95.A student rated P-A X>X is 20.73% more likely to score above the criterion average score than a student rated P-A X<X.

96.A student rated P-A X>X is 17.71% more likely to score below the criterion average score than a student rated P-A X<X.

103.A student rated P-A X>X is 19.97% more likely to score above the criterion average score than a student rated W X>X.

111.A student rated W X>X is 15.37% less likely to score above the criterion average score than a student rated E X>X.

119.For S.O.L.O. level M and R, a student rated Unsuccessful, W X>X, At X>X is 17.50% more likely to score below the criterion average score than a student rated Successful, W X>X, At X>X.

120.For S.O.L.O. level M and R, a student rated Successful, CC X>X, E X>X is 20.92% more likely to score above the criterion average score than a student rated Unsuccessful, CC X>X, E X>X.

121.For S.O.L.O. level M and R, a student rated Successful, E X>X is 52.00% more likely to score above the criterion average score than a student rated Unsuccessful, E X>X.

122.S students are 21.16% more likely to score above the criterion average score than U students (At X>X, Q-R X>X, E X>X).

125.S students are 15.85% less likely to score below the criterion average score than U students (P-A X>X, Q-R X>X).

126.S students with E X>X and CF X>X are 15.80% more likely to score above the criterion average score than below it.

127.S students with W X>X, At X>X, Q-R X>X are 16.57% less likely to

score above the criterion average score than below it.
128.U students with CF X>X and CC X>X are 30.00% more likely to score above the criterion average score than below it.
129.U students with E X>X, W X>X, At X>X, P-A X>X and Q-R X>X are 30.00% more likely to score below the criterion average score than above it.
130.Are above average S students different from all others on CF X>X ? Yes.
S X>X versus U X>X: W.N.P.D.=-11.93%
S X>X versus U X<X: W.N.P.D.= 22.58%
S X>X versus S X<X: W.N.P.D.= 18.29%
131.Are above average S students different from all others on Q-R X>X ? No.
S X>X versus U X>X: W.N.P.D.= 11.17%
S X>X versus U X<X: W.N.P.D.=- 3.31%
S X>X versus S X<X: W.N.P.D.=- 6.06%
132.Are above average S students different from all others on Q-R X<X ? No.
S X>X versus U X>X: W.N.P.D.=- 6.74%
S X>X versus U X<X: W.N.P.D.=-11.00%
S X>X versus S X<X: W.N.P.D.= 2.69%
133.Are above average S students different from all others on P-A X>X ? No.
S X>X versus U X>X: W.N.P.D.= 2.65%
S X>X versus U X<X: W.N.P.D.=-17.50%
S X>X versus S X<X: W.N.P.D.= 1.63%
134.Are above average S students different from all others on P-A X<X ? No.
S X>X versus U X>X: W.N.P.D.= 4.65%
S X>X versus U X<X: W.N.P.D.= 6.56%
S X>X versus S X<X: W.N.P.D.=- 1.63%
135.Are above average S students different from all others on W X>X ? Not all.
S X>X versus U X>X: W.N.P.D.= 5.53%
S X>X versus U X<X: W.N.P.D.=-13.92%
S X>X versus S X<X: W.N.P.D.=-16.00%
136.Are above average S students different from all others on CC X>X ? Not all.
S X>X versus U X>X: W.N.P.D.=-10.84%
S X>X versus U X<X: W.N.P.D.= 11.78%
S X>X versus S X<X: W.N.P.D.= 1.86%
137.Are above average S students different from all others on E X>X ? Yes.
S X>X versus U X>X: W.N.P.D.= 52.00%
S X>X versus U X<X: W.N.P.D.= 20.12%
S X>X versus S X<X: W.N.P.D.= 19.17%
138.Are above average U students different from all others on CF X>X ? Yes.
U X>X versus S X>X: W.N.P.D.= 11.93%
U X>X versus S X<X: W.N.P.D.= 31.15%
U X>X versus U X<X: W.N.P.D.= 35.55%
139.Are below average U students different from all others on CF X>X ? Not all.

U X<X versus S X>X: W.N.P.D.=-22.58%
U X<X versus S X<X: W.N.P.D.=- 7.58%
U X<X versus U X>X: W.N.P.D.=-35.33%
140.Are below average U students different from all others on Q-R X>X ? No.
U X<X versus S X>X: W.N.P.D.= 3.31%
U X<X versus S X<X: W.N.P.D.= 8.15%
U X<X versus U X>X: W.N.P.D.= 14.50%
141.Are above average U students different from all others on Q-R X>X ? Yes.
U X>X versus S X>X: W.N.P.D.=-11.17%
U X>X versus S X<X: W.N.P.D.=-19.00%
U X>X versus U X<X: W.N.P.D.=-14.50%
142.Are below average U students different from all others on the basis of At X>X ? Yes.
U X<X versus S X>X: W.N.P.D.= 18.29%
U X<X versus S X<X: W.N.P.D.=-40.00%
U X<X versus U X>X: W.N.P.D.=-11.32%
143.Are above average S students different from all others on the basis of At X>X ? Yes.
S X>X versus U X>X: W.N.P.D.= 31.50%
S X>X versus U X<X: W.N.P.D.= 11.32%
S X>X versus S X<X: W.N.P.D.=-27.89%
144.Is P-A X>X the same as CF X>X ? Maybe.
P-A X>X average percentage of all students=63.88%
CF X>X average percentage of all students=46.13%
P-A X>X standard deviation=15.33%
CF X>X standard deviation=18.61%
Correlation between percentage distributions of all P-A X>X and CF X>X students in all categories (S/U, above/below average, S.O.L.O. rating) is -0.29. Average inter-test correlations among the five classes separately is -0.24.
145.In which categories are the differences between above and below average students largest?
Q-R X>X: W.N.P.D.=- 8.88%
Q-R X<X: W.N.P.D.= 0.67%
P-A X>X: W.N.P.D.=- 4.35%
P-A X<X: W.N.P.D.=- 1.61%
W X>X: W.N.P.D.=-18.21%
CF X>X: W.N.P.D.= 24.34%
CC X>X: W.N.P.D.= 8.54%
At X>X: W.N.P.D.=-24.44%
E X>X: W.N.P.D.= 0.05%
That is, above average students are 8.88% less likely to score Q-R X>X than below average students, 24.34% more likely to score CF X>X, etcetera.

Significance testing of the trends in the W.N.P.D. table.

The percentages themselves cannot be examined for significance. One way of overcoming this problem is to group the W.N.P.D. values in some way and then examine if this grouping exhibits a significant trend. One example of this method is a test for significance using the Mann-Whitney U test.

A trend that was investigated, among many others, asked if there was a significant difference between Successful and Unsuccessful students who scored above average on criterion scores on the basis of above average Q-R score. There was a significant difference, $p<0.04$, with Successful students performing significantly better. The same was found for scientific Attitude, $p<0.04$, and for Enjoyment of science, $p<0.07$, but not for any other test categories. Similar significance tests were done on above average Successful to below average Successful students, idem for Unsuccessful students, comparisons on the basis of S.O.L.O. classifications, etcetera.

The example below investigated if above average Successful students were significantly different from above average Unsuccessful students on the basis of above average Q-R scores.

U sts.	Nu=2	S sts.	Ns=3	Calculations:
%	rank	%	rank	U1=Nu*Ns + Nu*(Nu+1)/2 -Sumu
25	5	44	3	=2*3 + 2*3/2 - 9
43	4	47	2	=0
		67	1	U2=2*3 + 3*4/2 - 6
Sumu= 9		Sums=6		=6

Ue=Nu*Ns/2= 2*3/2= 3
Standard deviation of U values = $\{Nu*Ns(Nu+Ns+1)/12\}^{0.5}$
= $\{2*3\ (2+3+1)/12\}^{0.5}$
= 1.73
Convert to standard Z-scores:
ZU1= (U1- Ue)/std.U dev.   ZU2= (6-3)/1.73
= (0-3)/1.73   = +1.73
= -1.73
Test for significance on the normal curve: sign. $p<0.04$.

Another interesting comparison, and similar ones like it, was the test for a significant relationship between the W.N.P.D. differences for Successful and Unsuccessful above average students and the W.N.P.D. values for students in general (S and U students together) scoring above and below the average criterion score on the bases of above average test category scores. The actual calculation is given next.

category	S vs. U X>X	sts X>X vs. sts X<X
Q-R X>X	11.17%	- 8.88%
P-A X>X	2.65%	- 4.35%
W X>X	5.53	- 6.39%
CF X>X	-11.93%	25.69%
CC X>X	-15.73%	8.54%
At X>X	31.50%	-24.44%
E X>X	52.00%	0.05%

Legend: S=successful
U=unsuccessful
X>X=above average
X<X=below average
vs.=versus

Correlation between the two columns: rho= -0.56
Significance test (Cohen, 1976, 332):
Standard deviation of correlation value $=(N-1)^{-0.5}$
$=(14-1)^{-0.5}$
$=0.28$
Equivalent t-score of correlation value=rho/std.corr.dev.
$t=-0.56/0.28$
$=-2.00$
Degrees of freedom= N-1= 14 - 1= 13
Correlation is significant, $p<0.035$.
If the E category is excluded, then rho=-0.90 at $p<0.006$.

Summary of significance calculations.

Many calculations like the example in the previous section were performed. A summary of these calculations is presented in this section. This summary should be read in conjunction with the W.N.P.D. data, because the significance values lend strength to the W.N.P.D. values.

In the tables below the descriptions:'better' and 'lower', in the third column of each table, refer to the average ranking of the designated group. For example, "Q-R X>X sign S better" means that the students who scored above average in the Q-R criterion differ significantly from U students with S students' average rank better, or higher, than the average rank of U students. That is, here S students outperformed U students.

Uabove av vs. Sabove av Sabove av vs. Ubelow av

Q-R X>X sign S better
Q-R X<X non-sign
P-A X>X non-sign
P-A X<X non-sign
W X>X non-sign
CF X>X non-sign
CC X>X non-sign
At X>X sign S better
E X>X sign S better

Q-R X>X non-sign
Q-R X<X non-sign
P-A X>X sign S lower
P-A X<X non-sign
W X>X sign S lower
CF X>X sign S better
CC X>X non-sign
At X>X non-sign
E X>X sign S better

(U+S)above av vs (U+S)below av

Q-R X>X non-sign
Q-R X<X non-sign
P-A X>X non-sign
P-A X<X non-sign
W X>X sign above av lower
CF X>X sign above av better
CC X>X sign above av better
At X>X sign above av lower
E X>X non-sign
E X>X sign when excluding Uabove av above av S better

Uabove av vs (U+S)above av, for S.O.L.O. M and R.

Q-R X>X non-sign
Q-R X<X non-sign
P-A X>X sign U better
P-A X<X non-sign
CF X>X sign U lower
CC X>X sign U lower
At X>X non-sign
E X>X non-sign

Ubelow av vs Sabove av, for S.O.L.O. M and R.

Q-R X>X non-sign
Q-R X<X non-sign
P-A X>X sign U better
P-A X<X non-sign
W X>X sign U better
CF X>X sign U lower
CC X>X non-sign
At X>X non-sign
E X>X sign U lower

Double comparisons in the W.N.P.D. data.

In table B.4 below a double comparison is undertaken. By this is meant a W.N.P.D. comparison is contrasted against another W.N.P.D. comparison. The question that was asked in the formulation of the first double comparison was: is there a significant relationship between above average S and U students on the one hand and students in general scoring above and below the average criterion score, on the basis of the categories Q-R X>X, P-A X>X, W, CF, CC, At, E ? The answer to this question was: yes, a significant correlation of -0.56 exists. This means that, essentially, comparing above average S and U students is the same as comparing above to below average students with the important distinction that the W.N.P.D. values will have the reverse signs attached to them. For example, Successful students scoring above average in Q-R are 10% more likely to score above the average than Unsuccessful students and, in general, students scoring above average in Q-R are 10% less likely to score above the average than below the average criterion score (note the negative correlation).

Table B.4 shows very high, significant relationships between the various double comparisons; especially the last double comparison is a virtual identity!

Table B.4 Correlation and significance data of various double comparisons of the W.N.P.D. data.

Categories being compared	test categories	rho	sign
1(S vs U above av) versus (above av vs below av)	Q-R X>X,P-A X>X W, CF, CC, At,E	-0.56	p<0.04
2(S vs U above av) versus (above av vs below av)	Q-R X>X,P-A X>X W, CF, CC, At.	-0.90	p<0.01
3(SX>X vs SX<X) versus (above av vs below av)	Q-R X>X,P-A X>X W, CF, CC, At,E	0.81	p<0.01
4(SX>X vs SX<X) versus (above av vs below av)	Q-R X>X,P-A X>X W, CF, CC, At.	0.93	p<0.01
5(SX>X vs SX<X) versus (above av vs below av)	Q-R X>X, W, CF, CC, At,E	0.95	p<0.01
6(UX>X vs UX<X) versus (above av vs below av)	Q-R X>X,P-A X>X W, CF, CC, At,E	0.76	p<0.01
7(UX>X vs UX<X) versus (above av vs below av)	Q-R X>X,P-A X>X W, CF, CC, At.	0.89	p<0.01
8(SX>X vs UX<X) versus (above av vs below av)	Q-R X>X,P-A X>X W, CF, CC, At,E	0.43	p<0.08
9(SX>X vs UX<X) versus (above av vs below av)	Q-R X>X,P-A X>X W, CF, CC, At.	0.45	p<0.08
10(SX>X vs UX<X) versus (above av vs below av)	Q-R X>X,P-A X>X W, CF, CC.	0.89	p<0.01
11(UX>X vs SX<X) versus (above av vs below av)	Q-R X>X,P-A X>X W, CF, CC, At,E	0.87	p<0.01
12(UX>X vs SX<X) versus (above av vs below av)	Q-R X>X,P-A X>X W, CF, CC, At.	0.95	p<0.01
13(UX>X vs SX<X) versus (above av vs below av)	Q-R X>X, W, CF, CC, At.	0.99	p<0.01

The tables below present the data that were used to derive table B.4.

Double comparison numbers 1 and 2.

	(S vs U above av)	(above av vs below av)
Q-R X>X	11.17%	- 8.88%
P-A X>X	2.65%	- 4.35%
W X>X	5.53%	- 6.39%
CF X>X	-11.93%	25.69%
CC X>X	-15.73%	8.54%
At X>X	31.50%	-24.44%
E X>X	52.00%	0.05%

Double comparison numbers 3, 4 and 5.

	(SX>X vs SX<X)	(above av vs below av)
Q-R X>X	- 6.06%	- 8.88%
P-A X>X	1.63%	- 4.35%
W X>X	-16.00%	- 6.39%
CF X>X	18.29%	25.69%
CC X>X	1.86%	8.54%
At X>X	-27.89%	-24.44%
E X>X	19.17%	0.05%

Double comparison numbers 6 and 7.

	(UX>X vs UX<X)	(above av vs below av)
Q-R X>X	-14.50%	- 8.88%
P-A X>X	-16.29%	- 4.35%
W X>X	-22.22%	- 6.39%
CF X>X	35.33%	25.69%
CC X>X	20.67%	8.54%
At X>X	-18.19%	-24.44%
E X>X	-34.50%	0.05%

Double comparison numbers 8, 9 and 10.

	(SX>X vs UX<X)	(above av vs below av)
Q-R X>X	- 3.31%	- 8.88%
P-A X>X	-17.50%	- 4.35%
W X>X	-13.92%	- 6.39%
CF X>X	22.58%	25.69%
CC X>X	11.78%	8.54%
At X>X	11.32%	-24.44%
E X>X	20.12%	0.05%

Double comparison numbers 11, 12 and 13.

	(UX>X vs SX<X)	(above av vs below av)
Q-R X>X	-19.00%	- 8.88%
P-A X>X	2.13%	- 4.35%
W X>X	-25.00%	- 6.39%
CF X>X	30.00%	25.69%
CC X>X	9.26%	8.54%
At X>X	-54.96%	-24.44%
E X>X	-38.52%	0.05%

APPENDIX C

AN EXAMPLE OF SIGNIFICANCE TESTING

The example below investigated if above average Successful students were significantly different from above average Unsuccessful students on the basis of above average Questioning-Recall scores.

U sts.	Nu=2	S sts.	Ns=3	Calculations:
%	rank	%	rank	U1=Nu*Ns + Nu*(Nu+1)/2 -Sumu
25	5	44	3	=2*3 + 2*3/2 - 9
43	4	47	2	=0
		67	1	U2=2*3 + 3*4/2 - 6
Sumu= 9		Sums=6		=6

Ue=Nu*Ns/2= 2*3/2= 3

Standard deviation of U values = {Nu*Ns(Nu+Ns+1)/12}0.5

= {2*3 (2+3+1)/12}0.5

= 1.73

Convert to standard Z-scores:

ZU1= (U1- Ue)/std.U dev.   ZU2 = (6-3)/1.73
   = (0-3)/1.73   = +1.73
   = -1.73

Test for significance on the normal curve: sign. p<0.04.

Another interesting comparison was the test for a significant relationship between the W.N.P.D. differences for Successful and Unsuccessful above mean students and the W.N.P.D. values for students in general (S and U students together) scoring above and below the mean criterion score on the bases of above mean instrument category scores. The calculation was performed as follows:

category	S vs. U (X>X)	sts X>X vs. sts X<X
Q-R X>X	11.17%	- 8.88%
P-A X>X	2.65%	- 4.35%
W X>X	5.53	- 6.39%
CF X>X	-11.93%	25.69%
CC X>X	-15.73%	8.54%
At X>X	31.50%	-24.44%
E X>X	52.00%	0.05%

Correlation between the two columns: rho= -0.56

Significance test (Cohen, 1976, 332):

Standard deviation of correlation value =(N-1)-0.5
=(14-1)-0.5
=0.28

Equivalent t-score of correlation value =rho/std.corr.dev.
t =-0.56/0.28
=-2.00

Degrees of freedom= N-1= 14 - 1= 13

Correlation is significant, p<0.035.

If the E category is excluded, then rho=-0.90 at p<0.006.

APPENDIX D

AN EXAMPLE OF AN F VALUE CALCULATION

MSerror	= SSerror/(ntotal-p);	ntotal	=number of students
	= 17,020.3/[(26+23)-2]	p	=number of groups
	= 326.13		

Sample	n-1	$(n-1)^{-1}$	s^2	$\log s^2$	$(n-1)\log s^2$
Te	25	0.04	335.10	2.53	63.25
Tc	22	0.05	392.82	2.59	57.07
	47	0.09			120.32

C $= 1 + (0.09-1/47)/3(2-1)= 1.02$

X^2 $= 2.303*C-1*[47(\log \text{MSerror}) - (n-1)\log s^2]$

$= 2.303*1.02-1*[47*\log 362.13 - 120.32]$

= - 0.12; at df=47 this value is non-significant and the ANOVA may continue.

Step one in the ANOVA calculates the total sum of squares: SS = 17237.80. The second step calculates the sum of squares between groups: SSbetween= 228.82. The third step calculates the sum of squares within the groups: SSwithin= 17008.98.

The total number of degrees of freedom= 48. The degrees of freedom between the groups= 1 and the degrees of freedom within the groups= 47. Therefore, the mean sum of squares betweeen the groups= 228.82/1= 228.82. The MS between groups= 17008.98/47= 361.89.

From these figures F= 228.82/361.89= 0.63. At the degress of freedom of 1 and 47, this F value is non-significant.

APPENDIX E

DETAILS OF THE DEVELOPMENT OF THE HIERARCHY OF INSTRUMENT CATEGORIES MODEL

1. A large negative correlation was found between Categorical-Functional and Whole, a similar negative correlation was also noted between Categorical-Class and Whole but not as large in absolute terms.
2. A large negative correlation exists between Enjoyment and Whole.
3. A large negative correlation between Application and Whole and a positive correlation between Principles-Application and Whole indicate that the negative correlation between Application and Whole is the cause of the positive correlation between Principles-Application and Whole.
4. There is a positive correlation between Categorical-Class and Categorical-Functional, especially for concrete students.
5. There is a positive correlation between Application and Categorical-Functional.
6. With the exception of the class designated Green Machine, three treatment classes had a large negative correlation between Principles-Application and Categorical-Functional. This is caused by a negative correlation between Principles and Categorical-Functional and a positive correlation between Application and Categorical-Functional.
7. There is a positive correlation between Attitude and Categorical-Class. This suggests that those students with an above average scientific Attitude tend to

score high in Categorical-Class. Therefore, students with above average Attitude and Categorical-Class scores should do well in inquiry based science courses, because there is a positive correlation between Categorical-Class and achievement score.
8. The data show a positive correlation between Application and Categorical-Class. In more detail, an above average Application score is an advantage for concrete students and an above average Categorical-Class score is an advantage for all students. Therefore, the positive correlation between Application and Categorical-Class is especially valid for concrete students.
9. A positive correlation between Questioning-Recall and Categorical-Class was found to be particularly strong for transitional and formal students. This is consistent with conclusion 3 in the C.C.P.I. section of this chapter.

The above nine conclusions are used in the next few paragraphs to develop a model.

10. Based on a negative correlation between Questioning-Recall and Categorical-Class, a large negative correlation between Questioning-Recall and Attitude and a positive correlation between Categorical-Class and Attitude, it can be concluded that Questioning-Recall measures something else than Attitude and Categorical-Class. it was noted previously (in the fourth step's conclusion 4) that only for formal students an above average Questioning-Recall score is an advantage and that for all other students it is a disadvantage.

This suggests a hierarchy: at first foster above average Attitude and Categorical-Class scores as a prerequisite to using above average Questioning-Recall scores as an advantage.

11. Because Attitude has a zero correlation with Principles-Application and for concrete students there is a negative correlation between Principles-Application and Categorical-Class, Attitude is not a prerequisite for Principles-Application (Attitude and Categorical-Class both apply mainly to concrete students).

For concrete students there is a negative correlation between Principles and Categorical-Class and a large positive correlation between Application and Categorical-Class; hence, for concrete students Categorical-Class should be encouraged through Application not Principles.

For transitional students there is a positive correlation between Principles and Categorical-Class and a zero correlation between Application and Categorical-Class; hence, for transitional and formal students while focussing on Application does not harm Categorical-Class should be encourage through Principles.Therefore, it can be concluded that Application is prerequisite for Categorical-Class.

12. A positive correlation between Application and Categorical-Functional, a negative correlation between Principles and Categorical-Functional, for three classes there was a large negative correlation between Principles-Application and Categorical-Functional and a positive correlation between Categorical-Functional and Categorical-Class for students other than formal (formal students have a zero correlation) all suggest that Categorical-Functional is largely identical to Application.

13. Enjoyment and Attitude have a large positive correlation especially for transitional and formal students, this suggests Enjoyment is a prerequisite for Attitude for the inquiry method espacially for transitional and formal students.

However, a large negative correlation between Questioning-Recall and Attitude for transitional and formal students suggests that scientific Attitude is fostered by concentrating on Recall not on Questioning. That is, Questions can only be asked if relationships are understood. In other words, a negative correlation between Attitude and Questioning-Recall must mean a positive correlation between Attitude and Recall.

There is no correlation between Enjoyment and Principles. There is a zero correlation between Attitude and Principles-Application. There is a small positive correlation between Attitude and Principles and a zero correlation between Attitude and Application. It could be hypethesized that perhaps Principles is prerequisite for Attitude. The correlation between Principles and Questioning-Recall is large and negative. Therefore, Principles must be positively correlated with Recall. Hence, Principles is a prerequisite for Recall, but not for Questioning. Attitude was also a prerequisite for Recall and this reinforces the hypothesis that Principles is prerequisite for Attitude.

The model that is starting to emerge from these conclusions suggests the following sequence:

```
P-----------
E-----------   At-------
                          R------QR
A-----------   CC------
CF---------
```

14. It seems strange that Principles and Application are prerequisites for different categories. The data, however, show that this is in fact the case: Principles has a large positive correlation with Principles-Application for concrete students. Principles has a negative correlation with Application for concrete students. Principles has a positive correlation with Application for transitional and formal students. Application has a large negative correlation

with Principles-Application. Principles has a negative correlation with Categorical-Functional. In all, for concrete students, Principles and Application are opposites.

Principles is large and negatively correlated with Questioning-Recall, therefore Principles is positively correlated with Recall. For concrete students Principles and Whole are somewhat positively correlated. For concrete students both Categorical-Functional and Categorical-Class are negatively correlated, but for transitional and formal students Categorical-Class and Principles are positively correlated and Principles and Whole have a zero correlation. These data would suggest that Whole is more appropriate to concrete students and Principles is more appropriate to formal students. Then, Whole must be prerequisite to Categorical-Functional, which in turn could be prerequisite to Principles for formal students.

A negative correlation exists between Categorical-Functional and Whole, especially for concrete students. There is also a negative correlation between Categorical-Class and Whole. Whole also has a negative correlation with Application, but a positive correlation with Principles-Application, at least for concrete students. Therefore, Whole must develop indirectly into Principles. Indirectly because Application is important for concrete students and Application is a prerequisite for Categorical-Class. Categorical-Class is negatively correlated with Principles for concrete students, but positively correlated for formal students. Hence, Principles is more appropriate for formal students not concrete students and Whole is more appropriate to concrete students. That is a very large transition and it is probable that Categorical-Functional and Categorical-Class must be intermediate steps. In fact, Categorical-Class is negatively correlated to Principles for concrete students but positive for formal students, indicating that Categorical-Class is more appropriate to transitional students than to concrete students.

The model developed earlier has now taken on a different appearance:

```
Part-Whole-----Whole-----CF&A-----CC------P------
                                                  R-----QR
                E----------At----------------------
 pre-concrete(?)    concrete  transitional   formal
```

Attitude is mainly for transitional and formal students because Principles was found to be for formal students, Attitude has positive correlations with Categorical-Class and Categorical-Functional, Attitude is negatively correlated to Questioning-Recall and, therefore positively correlated with Recall for formal students but much less for concrete students.

APPENDIX F

A list if teaching plans for all five treatment classes

The five classes will be dealt with in the order given below. A selection of tasks is presented in the order in which they were performed by the students.

(1) Chemistry in the Market place; grade ten; students mainly S.O.L.O. Relational and General Abstract.
(2) The Green Machine; grade ten; students ranged from S.O.L.O. Multiple to General Abstract.
(3) Electricity and Magnetism; grade ten advanced science; students only S.O.L.O. Relational and General Abstract.
(4) Bubbles and Bangs; grade nine; students mainly S.O.L.O. Singular and Multiple.
(5) Essential Chemistry; grade nine; students mainly S.O.L.O. Relational and General Abstract.

F.1 Chemistry in the Market place (grade ten)

1. Use one, some or all of the following materials to test if samples are acidic or basic (alkaline).
Material: phenolphtalien indicator (p-ind.), methyl-orange indicator, iron

filings, Mg strips, Al strips, pieces of Zn, copper oxide.
Samples: lemon juice, HCl (dil.), vinegar, sulphuric acid, sodium hydroxide, potassium hydroxide, sodium bicarbonate, apple, toothpaste, lime, milk, orange juice.

In the report include in your conclusion how to test for acids and how to test for bases.
Time: 4 periods.

2. Some acids are strong, some are weak. Using a base of given strength suggest a method to compare the strengths of citric acid and vinegar.
Material: p-ind., universal ind., burettes, 0.5 M NaOH, vinegar, solid citric acid.
Time: 3 periods.

3. Plan and carry out experiments to determine if the following compounds are ionic or covalent.
Equipment: water, power packs with volt- and ammeters.
Material: nitrates, carbonates, chlorates, chlorides, sugar, salt, alcohol, ammonia (dil.), kerosene, petrol, wax, fat, petroleum jelly, lubricating oil, pentane.
Time: 4 periods.

4. Compounds containing carbon are classified organic. Compounds that do not contain carbon are inorganic. In earlier days chemists thought that only living organisms contained carbon. Hence, carbon compounds were termed organic.
Question: what common gas contains carbon?
Question: if a substance is burned and gives off this gas what conclusion could you draw?
Teacher: explain lighted splint and bubbling through Bromothymol-blue techniques.

Students: classify the following chemicals:
kerosene, petrol, lube oil, pentane, alcohol, ammonia (NH_3), nitrates, carbonates, chlorides.
Time: 4 periods.

5. Compare your data in 3. and 4.
(i) Reorganize your data in a different way to the way you presented them in 3. and 4.
(ii) Are there patterns or trends in your data?
(iii)Draw one or more conclusions based on either the presence of any trends or the lack of any trends.
(iv) Ask a new question that you could investigate.
(v) Formulate a hypothesis from your new question.

6. Distinguish between the following chemical samples any way you can using only the materials provided.
Samples: several of each of cyclo compounds, alkanes, alkenes, alcohols;
Materials: water, acid, base, iodine crystals, methylated spirits (metho), petrol.
Note: (1) use small amounts;
(2) do not use flames;
(3) replace lids on bottles.
Time: 8 periods.

Examples of Assignments

1. Hydrogen peroxide is a liquid with strong bleaching power. It has a limited shelflife, that is it deteriorates quickly. It decomposes into water and oxygen.
The graphs below show how much oxygen is produced:
(a) when hydrogen peroxide decomposes naturally;
(b) when manganese dioxide is placed in the hydrogen peroxide (MnO_2

is a solid);

(c) when gold is placed in the hydrogen peroxide.

oxygen
produced
per
minute
(ml.)

time (min.) time (min.) time (min.)

Note: all three graphs used the same amount initial volume of hydrogen peroxide.

(i) How do the three graphs differ?

(ii) Can you draw any conclusions?

(iii) What do the three graphs have in common?

(iv) Does this contradict your conclusion in (ii)? If yes, revise your conclusion.

(v) Look at the table below. What conclusions can you draw?

Chemical name	mass of chemical before placing in hydrogen peroxide (grams)	mass of chemical after placing in hydrogen peroxide with the reaction completed (mass in grams)
manganese dioxide	5.01	5.01
gold	8.43	8.43

2. 10 g. of blue copper sulphate crystals, a pure substance, was dissolved in water. A positive and a negative electrode were placed in the solution and a

low voltage was applied. An ammeter showed a current flowed. Also, it was noticed that a reddish-orange deposit covered the negative electrode. Nothing was deposited on the positive electrode.

After several days, the reddish-orange deposit was quite thick and the solution was much less blue. The ammeter showed that the current was almost stopped even though the same voltage was still applied to the electrodes. The power was switched off and the negative electrode was weighed: the original electrode mass was 30 g., the final mass of the electrode was 33 g.

(a) How many grams of copper sulphate were dissolved?
(b) Was the copper sulphate pure?
(c) What was the reddish-orange deposit?
(d) Where did the deposit come from?
(e) How would you prove that the deposit did not come from the water?
(f) Why did the solution become less blue?
(g) Why was the current nearly stopped after several days?
(h) Why did one of the electrodes weigh more afterwards?
(i) List all the things you know about the copper sulphate solution (I can think of 13).
(j) The copper sulphate was pure, or was it not?
What seems to be the contradiction?
What could this suggest for some pure substances?

F.2 The Green Machine (grade ten)

1. Given the following classification scheme:

living things

animals plants micro-organisms

non-vascular vascular

non-vascular				vascular		
algae	lichen	fungi	mosses and liver-worts	cone bearing plants	flowering plants	ferns

Collect samples from outside and use the specimens provided to determine how you can distinguish between vascular and non-vascular plants.
Material: liquid dyes (2 colours), water, scalpels, wooden boards, microscopes, hand lenses.
Time: 5-6 periods.

2. Given: 7 commercially prepared microscope slides of vascular and non-vascular plants.
Compare these slides with the characteristics developed in 1. to distinguish between these two types.
Time: 1 period.

3. What gives plants their green colour?
Material: mortar and pestle, metho, filter paper, acetone, iodine, chalk powder, aluminium oxide, glass tubing, cotton wool.
Ask for anything else you want.
Time: 5-6 periods.

4. Plants cannot live without water. They need it to make food, to take up nutrients from the soil and it helps to give the plant rigidity.
Formulate a hypothesis about water and plants and test the hypothesis.
Material: water, petroleum jelly.
Time: 6 periods.

5. From the question: I wonder what material makes a permeable membrane?

Formulate a hypothesis and test it.

Material: balloon, filter paper, glad wrap, silver foil, grease proof paper, ordinary paper, photo copying paper, cellulose tubing.

Time: 2 periods.

6. Design and carry out an experiment to investigate how plants respond to one or more of the following:

light (phototropism);

gravity (geotropism);

water (hydrotropism);

touch (haptotropism);

<u>Examples of assignments</u>

Photosynthesis

1. What would you need to remove from air, if you wanted to find out if plants produce oxygen in sunlight?
2. What would you need to do to the air and to the plant if you wanted to find out if it is the sunlight that helps to produce oxygen?
3. How did you prove that plants produce starch in the presence of sunlight?
4. Sodium hyposulfite is a liquid that can remove oxygen from the air if the air is bubbled through it.

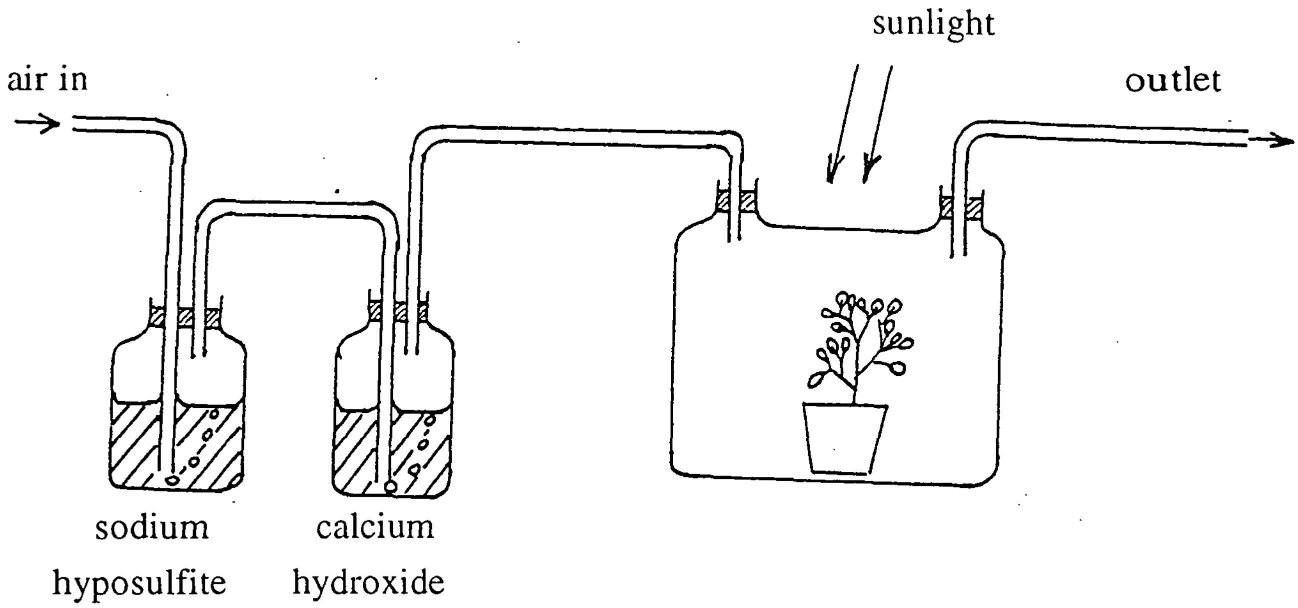
Calcium hydroxide is a liquid that can remove carbon dioxide from the air if the air is bubbled through it.

How would you prove that carbon dioxide is necessary to a plant if it is to produce oxygen and starch in the presence of sunlight?

5. Bromothymol-blue is normally a blue liquid, but when CO_2 is bubbled through it, it turns yellow.

In question 4., how would you prove that all the CO_2 has been removed from the air before pumping the air to the plant?

6. What would the following experiment prove?



Method: An air pump pushes air through the two bottles. The plant was removed from the classroom and placed in the large jar when the air started to flow through.

Results: (i) Iodine confirmed starch is present in the leaves after several hours.
(ii) at the outlet the air was bubbled through Bromothymol-blue and it stayed blue.
(iii) at the outlet the air was tested for oxygen and no oxygen was found.

7. How could the experiment in 6. be improved?

F.3 Electricity and magnetism (Grade ten Advanced science)

1. Devise a test for measuring magnetic force quantitatively.
Material: bar magnets, magnetic compasses, string, force balances, wire, power pack, iron nails, iron filings, clear plastic sheets.
Time: 4-5 periods.

2. Use the test devised in 1. to measure if the magnetic force of weak magnets can be improved.
Time: 2 periods.

3. Magnetic field lines are imaginary lines with arrows that indicate the direction in which a compass needle would be forced if the needle were placed on that spot.
(i) Investigate magnetic field lines for a single magnet.
(ii) As (i) for 2 magnets.
(iii) How do strong magnets compare to weak ones?
Time: 0.5 period.

4. The word 'flux' refers to a collection of parallel field lines. Flux density (B) is a measure of the number of magnetic field lines per area.
(i) Investigate B for strong and weak magnets.
(ii) As (i) for a coil with different d.c.. currents going through the coil.
Time: 0.5 period.

5. Investigate B for coils of different lengths but have the same wire diameter and coil diameter.
Time: 1 period.

6. Use a thick, single, straight wire to investigate the relationship between the direction of current and the magnetic flux around the wire.

7. Investigate the force on a wire through which a current flows and is held in a magnetic field.

8. Use the equipment provided to find a relationship between the current through a loop and magnetic field.
Equipment: 2 solid copper rings of approximately 10 cm. diameter fixed at 900 to each other, but sharing a common axis of symmetry in the plane of the

rings. A compass is placed on cardboard in the centre of the two rings.

(For 6,7 and 8 all the necessary equipment was set up for the students in different corners of the laboratory).
Time: for 6,7 and 8 together 2 periods.

9. Investigate the effect on the magnetic field lines when an object is placed between two magnets lying end to end.
Material: annular ring, solid circle, rectangular bar, bar with two convex opposite sides, bar with two concave opposite sides. Different materials for most shapes: Al, Fe, Cu, plastic.

<u>Example of a worksheet on alternative explanations</u>

This worksheet was given to students to demonstrate the need for thoughtful planning during the methodology phase of experimentation. The worksheet illustrates the necessity of controlling all relevant variables.

Below are given experimental data and correct, or partly correct, conclusions. For each suggest an alternative, correct conclusion.
1. To prove that magnetic fields act in 3-D, a scientist filled a small beaker with glycerol (a syrupy liquid more dense than water) and mixed in some iron filings till they were evenly distributed. She then suspended a small magnet in the beaker but observed no change in the iron filings.
Conclusion: magnetic fields do not act in 3-D.
Alternative:
2. To prove that sparks demagnetize magnets, a scientist struck a spark between a magnet and a soft, non-iron electrode at 400V. He measured the magnetic force before and after and found a reduction in strength had occurred.
Conclusion: sparks demagnetize magnets.
Alternative:

3. More 15-year old students from poor families than from rich families get summer jobs. On the other hand, more 15-year old students from rich families travel overseas or engage in some other educational endeavour.
Conclusion: the wealthy care more about educating their children than do the poor.
Alternative:
4. In each of the above suggest an experiment to conclusively reject or accept the conclusions provided.

<u>Examples of assignments</u>

1. (a) Draw the magnetic field around one magnet.
   (b) What do the lines represent?
   (c) What do you have to place in the field to determine the field's direction and strength?
   (d) Why are the lines bent?

2. (a) Look at figure 1(a); what do the lines represent?
   (b) Look at figure 1(b); why is the gravitational field bent around the moon?
   (c) What do you have to place in the field to determine the field's direction and strength?

3. (a) Look at figure 2(a); what do the lines represent?
   (b) How are the lines in figures 2(a)i and 2(a)ii different?
   (c) How is this similar to a magnet?
   (d) Look at figures 2(b) and 2(c); why are the lines bent?
   (e) What do you have to place in the field to determine the field's direction and strength?

4. In separate experiments the following data were collected:

GRAVITATIONAL FIELD			ELECTRIC FIELD		
force on mass (N)	grav. field strength ($N.kg^{-1}$)	mass placed in field (kg)	force on charged particle (N)	electric field strength ($N.C^{-1}$)	charge of particle placed in ($C*10^{-19}$)
12.5	10.0	1.250	$-4.8*10^{-14}$	$3.0*10^{5}$	-1.6021
26.0	130.0	0.200	$+3.8*10^{-15}$	$2.4*10^{4}$	+1.6021
1.0	10.0	0.100	$+3.2*10^{-13}$	$1.0*10^{6}$	+3.2042
1.2	0.1	12.350			

Note: C=Coulomb, N=Newton.

(a) For each of the fields separately can you see any trends?

(b) Derive a mathematical equation for each field.

(c) Can you infer any relationships from these trends?

(d) Are there any similarities between questions 1,2,3 and the data in question 4?

(e) Based on your answers to (b), (c) and (d), hypothesize an equation that would apply to a magnetic field. Identify all variables and explain the relationship.

5. If you have enjoyed this assignment, even though it was difficult, please continue:

(a) Find out what Lorentz Law is and explain the variables and their relationships (it ties together all the work you have done in this assignment).

(b) Look at the charges of the particles in the Electric field data in question 4. Find out what particles they were.

(c) How could the data in question 4 have been obtained?

(d) How could you obtain similar data for a magnetic field?

(e) Suggest a title for this assignment.

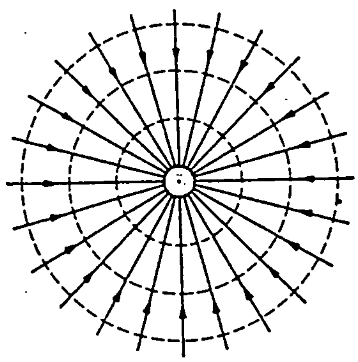


Fig. 1(a) The solid lines represent the gravitational field around a mass.

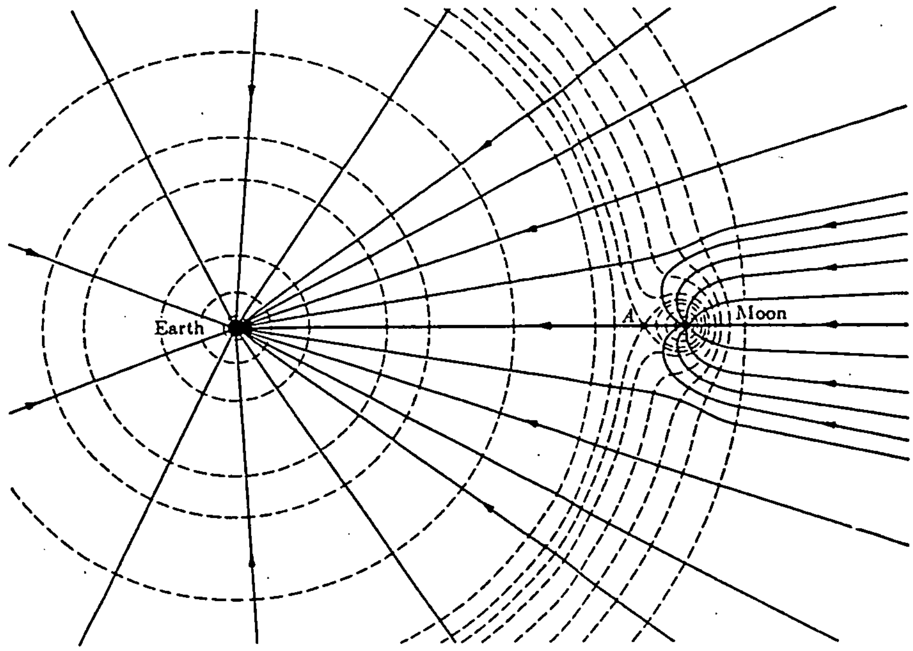


Fig.1(b) The solid lines represent the total gravitational field around a large mass and a nearby smaller mass.

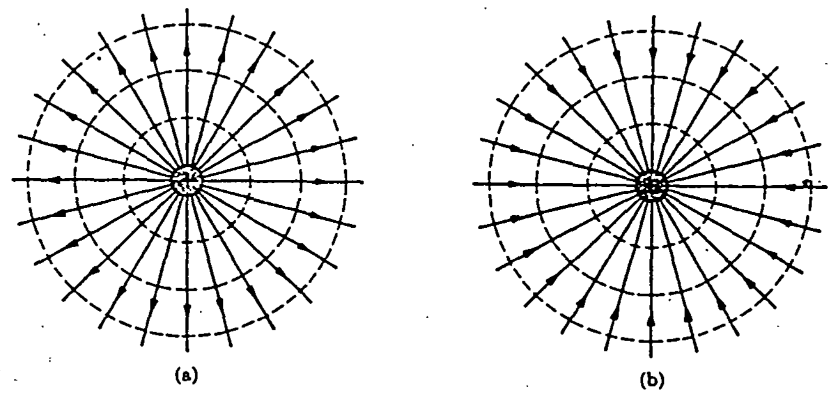
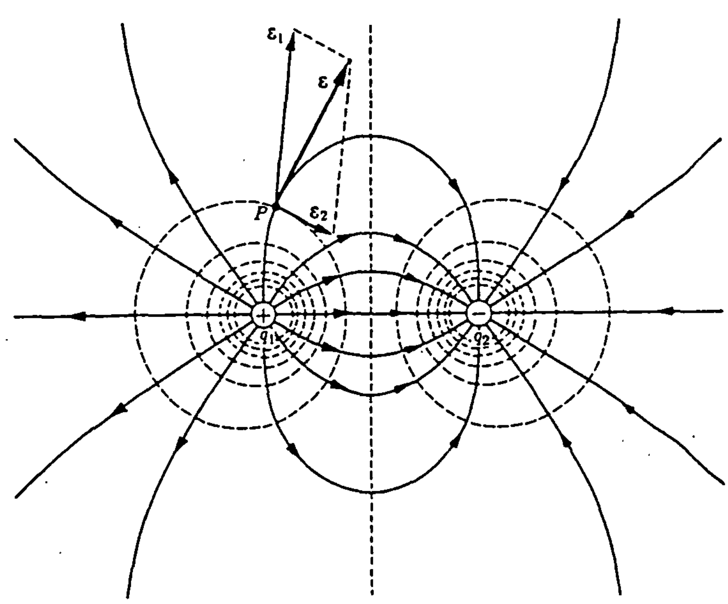


Fig.2(a) The solid lines represent the electric fields around charged particles.

Fig.2(b) The solid lines represent the total electric field around two oppositely charged particles.



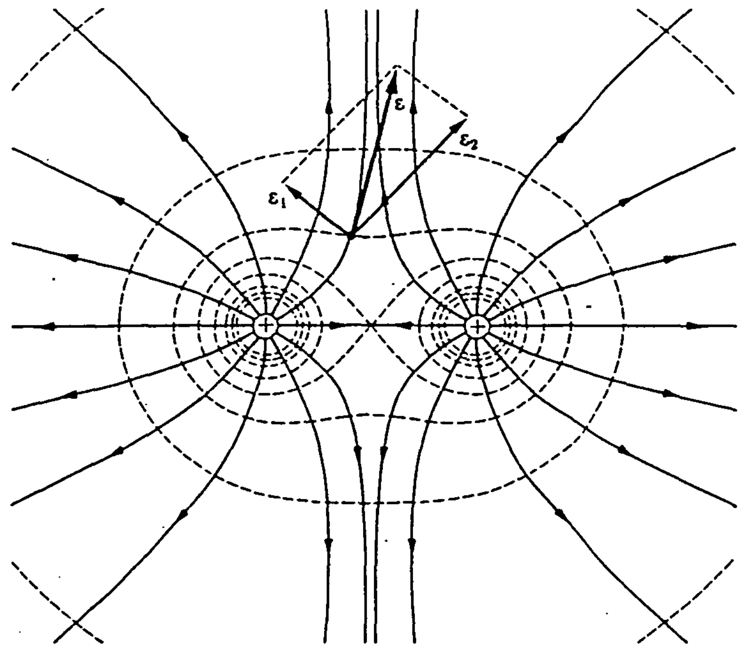


Fig.2(c) The solid lines represent the total electric field around two positively charged particles.

F.4 Bubbles and Bangs (Grade nine)

1. Determine the boiling points of the following liquids:
water, alcohol, glycerol, olive oil, hexane (fume cupboard!)
Question: what is a boiling point?
Equipment: Pasteur flasks.
Time: 3 periods.

2. How are the boiling points affected when the liquids are mixed?
Time: 2 periods.

3. From 2. write two questions. Formulate a hypothesis from each question and test one of the hypotheses.
Time: 2 periods.

4. A mixture is distinguished from a pure substance if its ingredients can be separated. The separating techniques we will use are: filtering, evaporating,

draining, centrifuging, pipetting, decanting, chromatography, preferential dissolving, looking under the microscope.

(i) Is distilled water pure?

(ii) Is tap water pure?

(iii) Make the following mixtures. Then, prove they are mixtures by separating them again:

Sugar and water;

Salt and water;

Sugar and sulphur;

Sugar and salt.

(iv) Is petrol a mixture?

(v) Make the following mixtures. Then, prove they are mixtures by separating them again: petrol and water; alcohol and water.

Time: 6 periods.

5. Teacher demonstration of instruments for the presence of the gasses: oxygen, hydrogen, carbon dioxide, water.

When some substances are heated a gas evolves. Was the substance really pure?

Materials: carbonates, chlorates, sodium nitrate, potassium nitrate, water, liquid and solid sodium bicarbonate, iron, copper.

Time: 4 periods.

6. Which substances lose weight upon heating?

Materials: carbonate of Mg, Zn, Fe, Cu; sodium nitrate, potassium nitrate, Al, Mg, Zn, Fe, Cu, graphite, sulphur (in fume cupboard!)

Were the substances pure or were they mixtures?

Equipment: electronic balance (three decimal places).

7. What can you find out about the electrolysis of carbonates (Zn, Cu, Fe, Pb),

chlorides (Zn, Cu, Fe, Pb), Zn, Cu, Fe, S.
Material: power packs, clean nails, emery paper, electronic balance.
Time: 4-5 periods.

8. Do all chemical reactions take place at the same rate?
Material: Al, Mg, Zn, acids of different concentrations, thermometers, clear plastic tubes, hot and cold water.
Time: 3 periods.

9. What are the differences between acids and bases?
Material: acids and bases of several different concentrations, p-ind., universal-ind., litmus paper, Mg, Zn, Cu, MgO, ZnO, CuO.
Time: 3 periods.

F.5 Essential Chemistry (Grade nine)

1. Do at least two physical changes and two chemical changes. Report on why they are physical or chemical changes.
Material: Fe filings, powdered S, magnet wrapped in tissue, olive oil, salt, methylated spirits, HCl (dil), vinegar, sodium bicarbonate, iodine (solid).
Time: 2 periods.

2. Do experiments to prove that the substance is pure or a mixture. You can use any of the following methods: filtering, evaporating, draining, centrifuging, pipetting, decanting, chromatography, preferential dissolving, looking under the microscope.
Material: salt, sugar, metho, water, milk, orange juice, Fe and S mixture in 1:1 proportion by mass both before heating and after heating (in fume cupboard!), margarine.

3. Teacher demonstration of tests for the presence of the gasses: oxygen, hydrogen, carbon dioxide, water.

When some substances are heated a gas evolves. Was the substance really pure?
Materials: carbonates, chlorates, sodium nitrate, potassium nitrate, water, liquid and solid sodium bicarbonate, iron, copper.
Time: 4 periods.

4. Which substances lose weight upon heating?
Materials: carbonate of Mg, Zn, Fe, Cu; sodium nitrate, potassium nitrate, Al, Mg, Zn, Fe, Cu, graphite, sulphur (in fume cupboard!)

Were the substances pure or were they mixtures?
Equipment: electronic balance (three decimal places).

5. What can you find out about the electrolysis of carbonates (Zn, Cu, Fe, Pb), chlorides (Zn, Cu, Fe, Pb), Zn, Cu, Fe, S.
Material: power packs, clean nails, emery paper, electronic balance.
Time: 4-5 periods.

6. Do reactions with the following pure substances. Conclude whether you have ended up with a new compound or a mixture. Record all observations.
(a) Fe fine filings + powdered S in the ratio 1:1 by mass. Heat in fume cupboard.
(b) $CaCO_3$ (calcium carbonate) + HCl, add HCl drop by drop.
(c) CaO (calcium oxide) + H_2O, add H_2O drop by drop.
(d) $Pb(NO_3)_2$ (lead nitrate) + KI drop by drop.
(e) Mg + HCl, place a small strip of Mg in a little HCl.
(f) Zn + I_2 (solid). Powder the zinc using a file. Crush the iodine crystals in a mortar and pestle. Place equal amounts (dry) in a test tube. Add a little H_2O drop by drop. It may need a little heat.
Questions to be answered in your report:

(i) were the original substances pure?

(ii) if you ended up with a suspension or precipitate, you may need to separate the liquid from the solid and determine if the solid is pure.

(iii) how can you tell the new substance is different from the original?

Time: 4-5 periods.

7. Which of the elements below are metals? Which are non-metals? Why are they metals or non-metals?

Material: water, acids, bases, battery, ammeter, several leads, universal ind., several metals and non-metals, both to experiment with and to observe only, for instance Hg.

Teacher demonstrations on Na, Li, P, etc., with water and tested afterwards with universal ind.

As always, present your data in several different ways, find patterns and trends, draw conclusions. Then, ask new questions.

Time: 3 periods.

8. Design experiments to tell the differences between Fe, Cu, Mg, Zn, FeO, CuO, MgO, ZnO.

Use everything you have learned so far: physical and chemical change, pure substances and mixtures, making compounds, acids, bases, pH.

Remember to ask questions, formulate hypotheses, observe, record, patterns?, specific conclusion?, general conclusion?, new questions and new hypotheses.

Time: 4-5 periods.

9. Some acids are strong, some are weak. Using a base of given strength suggest a method of comparing the strengths of citric acid and vinegar.

Material: p-ind., universal ind., burettes, 0.5 M NaOH, vinegar, solid citric acid.

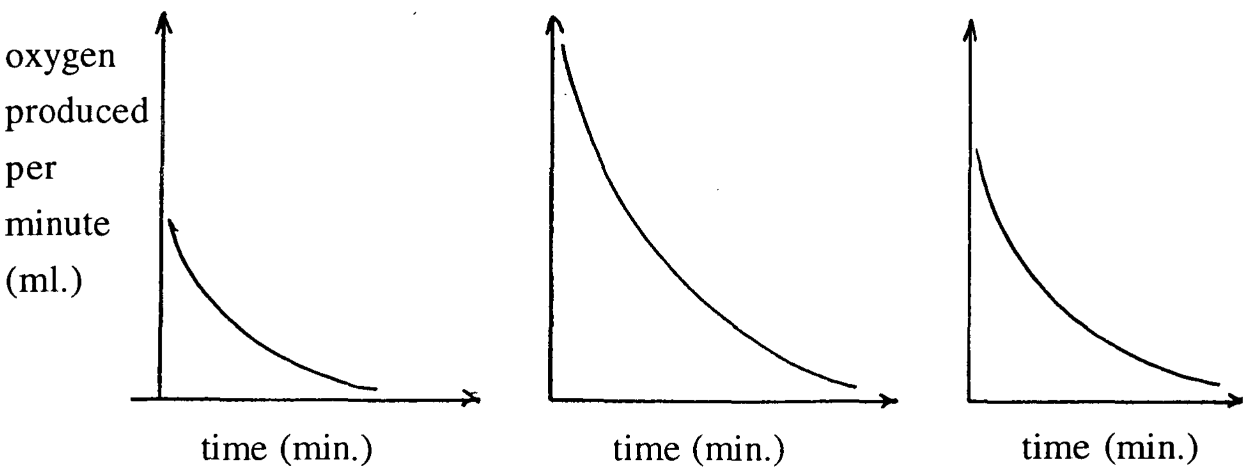
Time: 3 periods.

10. Do experiments to determine if the following compounds are ionic or covalent: nitrates, carbonates, chlorates, chlorides, sugar, salt, alcohol, ammonia (dil.), kerosene, petrol, wax, fat, petroleum jelly, lubricating oil, pentane.
equipment: water, power packs with volt- and ammeters.
time: 4 periods.

11. Are acids and bases ionic or covalent?
Material: several acids and bases both solid and liquid, ask for anything else you want.

<u>Examples of Assignments</u>

1. Hydrogen peroxide is a liquid with strong bleaching power. It has a limited shelf life, that is it deteriorates quickly. It decomposes into water and oxygen.
The graphs below show how much oxygen is produced:
(a) when hydrogen peroxide decomposes naturally;
(b) when manganese dioxide is placed in the hydrogen peroxide (MnO_2 is a solid);
(c) when gold is placed in the hydrogen peroxide.



Note: all three graphs used the same amount initial volume of hydrogen peroxide.

Q(a) How do the three graphs differ?

Q(b) Can you draw any conclusions?

Q(c) What do the three graphs have in common?

Q(d) Does this contradict your conclusion in Q(b)? If yes, revise your conclusion in question Q(b).

Q(e) Look at the table below. Can you draw any conclusions?

chemical name	mass of chemical before placing in hydrogen peroxide (grams)	mass of chemical after placing in hydrogen peroxide with the reaction completed (mass in grams)
manganese dioxide	5.01	5.01
gold	8.43	8.43

2. 10 g. of blue copper sulphate crystals, a pure substance, was dissolved in water. A positive and a negative electrode were placed in the solution and a low voltage was applied.

An ammeter showed a current flowed. Also, it was noticed that a reddish-orange deposit covered the negative electrode. Nothing was deposited on the positive electrode.

After several days, the reddish-orange deposit was quite thick and the solution was much less blue. The ammeter showed that the current was almost stopped eventhough the same voltage was still applied to the electrodes.

The power was switched off and the negative electrode was weighed: the original electrode mass was 30 g., the final mass of the electrode was 33 g.

(a) How many grams of copper sulphate were dissolved?

(b) was the copper sulphate pure?

(c) What was the reddish-orange deposit?

(d) Where did the deposit come from?

(e) How would you prove that the deposit did not come from the water?

(f) Why did the solution become less blue?

(g) Why was the current nearly stopped after several days?

(h) Why did one of the electrodes weigh more afterwards?

(i) List all the things you know about the copper sulphate solution (I can think of 13).

(j) The copper sulphate was pure, or was it not?
What seems to be the contradiction?
What could this suggest for some pure substances?