1990

# Secondary students' difficulties with designing a controlled experiment

Bradley E. Watson
*Edith Cowan University*

# SECONDARY STUDENTS' DIFFICULTIES WITH DESIGNING A CONTROLLED EXPERIMENT

by

Bradley Earl Watson

Dip. Tch. (W.A.C.A.E.)

A thesis submitted in partial fulfilment of the requirements for the award of Bachelor of Education with Honours in the School of Education, Western Australian College of Advanced Education.

Date of submission: June, 1990.

## ABSTRACT

This research investigation was conducted to identify the difficulties encountered by secondary science students in the design of a controlled experiment. These difficulties were illuminated by studying the combination of knowledge, skills and strategies employed by experts.

Three categories of individuals were used in the study, Year 10 science students, Year 12 science students and science lecturers who had completed doctoral studies in their fields. There were six subjects interviewed in each category.

Each of the subjects were given a task which involved thinking aloud while planning a controlled experiment. When the subjects had completed the task, they were asked several questions to probe their understanding of the various concepts involved in experimental design. Each interview was tape recorded, transcribed and analyzed.

The investigation revealed that the students generally had a poor understanding of many of the process skills used in planning experiments. In particular, the students experienced difficulty in hypothesizing, and in the identification and the control of variables.

Implications for instruction have been identified, as well as implications for further research.

I certify that this thesis does not incorporate, without acknowledgement, any material previously submitted for a degree or diploma in any institution of higher education and that, to the best of my knowledge and belief, it does not contain any material previously published or written by another person except where due reference is made in the text.

[redacted] Watson

## ACKNOWLEDGEMENTS

The author would like to express his sincere appreciation to his thesis supervisor Mr Mark Hackling, for his availability, unfailing patience, and his continual support and encouragement.

Thanks are also extended to the students and teachers at Corpus Christi College, and to the science lecturers at the Western Australian College of Advanced Education, for voluntarily taking part in the study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

# CHAPTER 1

## Introduction

Current science syllabi place a heavy emphasis on the development of those process skills necessary for conducting scientific investigations. The skill of experimental design is central to scientific inquiry and is therefore included in our state's Lower Secondary Science Syllabus (Education Department of Western Australia, 1981).

### Problem Statement

Research in Western Australia indicates that many secondary school students experience difficulty with designing a controlled experiment (Garnett, Tobin & Swingler, 1985; Tobin & Capie, 1980). Designing a controlled experiment is one of the more difficult of the integrated science process skills, as it incorporates a large number of lower and high level skills, including the ability to control variables. A chief aim of science education should be to promote the development of this ability, as it is essential for any form of scientific investigation.

### Rationale

Student difficulties in the area of experimental design need to be diagnosed so that appropriate remedial strategies can be developed, and recommendations for curriculum reform can be prepared. It is important to study how experts approach the problem, so that the

combination of knowledge, skills and strategies that give optimum performance can be identified. This information will give an indication of how current patterns of science instruction could be modified.

## Purpose and Research Questions

The purpose of this study is to identify the difficulties encountered by secondary science students in the design of a controlled experiment. These difficulties will be illuminated by studying the knowledge, skills and strategies used by experts in experimental design.

The study was conducted within the expert-novice paradigm, drawing on theoretical frameworks from the cognitive development, information processing psychology and planning literature. Two research questions were addressed by this investigation.

1. What difficulties are encountered by secondary science students in planning a controlled experiment?
2. What combination of knowledge, skills, and strategies is employed by experts in planning a controlled experiment?

# CHAPTER 2

## Literature Review

This chapter reviews the literature which pertains to an investigation of secondary students' difficulties with designing a controlled experiment. The review provides a theoretical framework to guide the design of the study, and provides a basis for the analysis and interpretation of results.

The purpose of the study is to identify difficulties that secondary students encounter when planning a controlled experiment. It is therefore necessary to consider the cognitive processes that take place during such an activity. The review initially considers theories of information processing and the development of expertise. Theories of planning are discussed because identification of various approaches to the "planning" of an experiment, is the chief goal of the project. Developmental differences associated with the ability to isolate and control variables are described, and studies of training for this ability are reviewed.

### Information Processing Psychology

It is a generally accepted view in cognitive psychology that ability at a certain task reflects both a person's domain specific knowledge and a set of skills for processing information. The thrust of research in cognitive psychology has been directed at studying how information is stored in memory, how this information may

be transformed, and how it is retrieved for use in further learning and problem solving (Stewart & Atkin, 1982).

Research indicates that there are three related but distinct memory stores: sensory information store, short-term memory and long-term memory (Lindsay & Norman, 1977). There are also several processes which facilitate the flow of information between these stores. The interaction between these stores and the transfer processes are shown in Figure 1.

The sensory information store (SIS) maintains detailed sensory information for approximately 0.1 to 0.5 seconds, and is a complete and accurate replication of environmental input. Short-term memory (STM) is capable of holding about eight pieces of information for a few seconds up to a few minutes (Wingfield, 1972). According to Stewart and Atkin (1982), STM has two functionally different aspects, an "echo-box", in which information is lost if not rehearsed, and a "working-memory" which is responsible for the information processing which occurs during problem solving. These aspects are shown in Figure 1.

Long-term memory (LTM) has a virtually unlimited storage capacity, and receives information in an organized fashion from STM. Information can be retrieved from LTM and worked upon in STM, so long as the system of retrieval matches the system of storage (Novak, 1980). STM uses concepts retrieved from LTM to interpret new information from SIS and evaluate its possible usefulness.

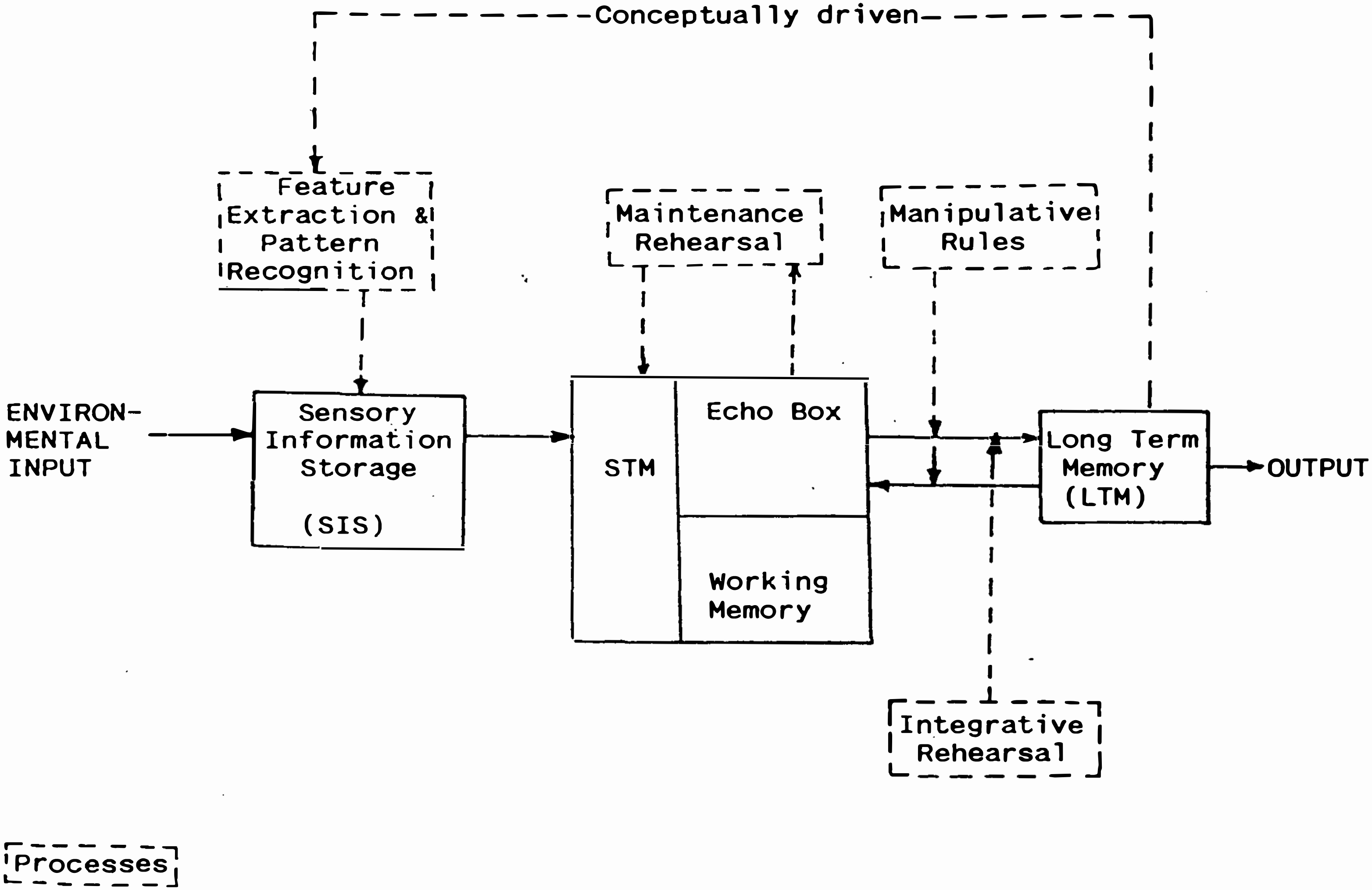


Figure 1. An Outline of Human Memory (Lindsay & Norman 1977)

Estes (1982) suggested that for efficient problem solving, two aspects of LTM are called upon, the episodic memory and the semantic memory:

> ... episodic memory as the source of analogies between a current problem and others that the person has solved, and semantic memory as the store of information concerning concepts and relationships relevant to a current problem. (Estes 1982, p. 214)

In LTM, both episodic and semantic information is stored in associative networks. Networks consist of nodes and links, "the nodes are units representing concepts, and the links represent relations between the concepts" (Cohen, 1983, p. 28). It is the interrelation between the nodes that forms the basis of information retrieval.

Stewart and Atkin (1982) have referred to the "monitor" and the "interpreter", as the two processes that store new information and retrieve old information. The monitor indicates to the interpreter that there is information in LTM that needs to be accessed, "then directs the interpreter to the appropriate area" (Stewart & Atkin, 1982, p. 326).

Knowledge related to a particular phenomenon is stored together in an organized body called a schema. The knowledge stored in the various schemas is processed by automatic and controlled processes. Controlled processes require focused attention and take effort to maintain, such as tackling a novel and challenging problem, whereas automatic processes occur with little or no conscious effort, walking and talking are examples of automatic

processes. Controlled processes maintain current goals, direct thought and behaviour, and take up a large proportion of working memory. In comparison "automatic processes make very small or possibly no demands on working memory, so their capacity is large relative to controlled processes" (Stillings et al., 1987, p. 52).

It is thought that the more complex cognitive, perceptual and motor skills require a combination of automatic and controlled processes:

> Controlled processes are used to maintain goals and flexibly direct skilled performance, to meet novel situations that could not previously have been automated. The lower level, more consistent components of the skill are performed automatically. (Stillings, 1987, p. 56)

In his review of the research pertaining to skill acquisition, Stillings et al. (1987) cited several researchers (Fitts, 1964; La Berge, 1976; Anderson, 1983) who found that skill acquisition is characterized by a transition from dominance of controlled to dominance of automatic processes. After extensive research on the development of automization, Shiffrin and Dumais (1981) argued that:

> Automatic processing will develop as skill acquisition proceeds. We think that automization is a major component of skill acquisition, in both the cognitive and motor domains, and suggest this factor be given prominent attention in research in these domains. (Shiffrin & Dumais, 1981, p. 139)

Automation is but one of many distinguishing characteristics of expertise at a particular skill, the features of expertise have been documented by several authors, and the range of these features is now discussed.

## The Development of Expertise

It is important to study the characteristics that distinguish experts from novices, as these characteristics suggest something about how instruction can be designed to develop expertise in any particular area (Bereiter & Scardamalia, 1986).

Hackling and Lawrence (1988) compared expert and novice use of genetics knowledge while solving genetics problems. They found that experts differed in their problem solving strategy, and in general completeness and conclusiveness of their solutions.

Experts tend to have an extremely refined system of retrieval from LTM. Chase and Ericsson (1981) found that retrieval systems of experts differed in the following ways:

> Firstly, experts seem invariably to know when to apply knowledge in a given task, whereas it is characteristic of novices that they often fail to apply what they know. Secondly, during the performance of some skilled task, experts store intermediate knowledge states for future reference in directly accessible locations in the long-term memory. (Chase & Ericsson, 1981, p. 175)

Thus experts have very well developed semantic networks, which facilitate the access of information relevant to the task at hand.

McGaw and Lawrence (1984) have suggested that experts differ from novices in the way they represent problems, by the forms of their solutions, and the control they have

over their own processes.

The content of schemas determines the mental model that an individual uses to represent a problem. Novices build familiar models with familiar objects, whereas an expert's model is typically more abstract and can be transformed in different ways to represent different solution paths (McGaw & Lawrence, 1988). Thus experts perceive and represent problems in a more sophisticated and abstract way, which enables them to select more effective problem solving strategies.

Another characteristic difference between expert and novice solution strategies, is what McGaw and Lawrence (1988) referred to as "the direction of processing". Novice solutions are characterized by a backward-working, means-end strategy, while experts employ a forward-working approach. McGaw and Lawrence exemplified this by reference to tasks involving a chain of steps linking givens to an unknown end:

> The expert forward-working strategy flows directly from appropriate problem representations. Experts start with the givens and proceed through the chain, novices first construct the chain by working backwards from the unknown end. (McGaw & Lawrence, 1988, p. 4)

The solution processes employed by experts are largely forward-working, flexible, and automated. McGaw and Lawrence defined yet a further distinguishing characteristic of experts, that is their control of processing. They argued that the use of automated processes for the routine aspects of problems, makes

available working memory capacity for setting goals, planning, and monitoring progress. This allows experts to have greater cognitive control over the planning of their problem solutions. Thus it appears that planning technique is a key discriminator between experts and novices, it is this aspect of human cognition which is now discussed.

## Planning

Hayes-Roth and Hayes-Roth (1979) proposed an opportunistic model of planning, the assumption is that at each point in the planning process, the planner's current observations and decisions, give rise to various new opportunities for plan development. Previous models assumed that planning involved a sequence of successive refinements on an inclusive, loosely specified initial plan. Research by Hayes-Roth and Hayes-Roth has contradicted this assumption, and provided evidence for their multidirectional model. That is, they found planning to be largely an event-driven, rather than a goal-directed behaviour.

Lawrence, Dodds and Volet (1983) compared the planning strategies of busy adult women and bright adolescent girls. The adults were found to be more efficient in their planning, and generally worked at a more abstract level. The adolescents tended to concentrate on immediate tasks and showed apparently random behaviour.

A model specific to the task of planning and

conducting a scientific investigation was proposed by Tobin and Capie (1980). The model illustrates the various cognitive skills necessary for experimental design: identifying and controlling variables, defining operationally, graphing and interpreting data (Tobin & Capie, 1980, p. 592).

The model, which is shown in Figure 2, provides a framework for planning and conducting a scientific investigation. The authors argued that if the skills are practiced within a variety of contexts, students can generalize their use to other content areas.

An individual's level of cognitive development, influences his/her ability to apply the integrated science process skills. Tobin and Capie (1980) found that 30 percent of variation in process skill achievement can be attributed to the formal reasoning capacity of the learners. As cognitive development has such a large influence on the ability to apply process skills, it is discussed at length in the next section.

1. State the problem.

2. Identify the dependent variable in the investigation.

3. Identify procedures for measuring the dependent variable in the investigation.

4. Identify variables that may affect the dependent variable in the investigation.

5. State hypotheses that can be tested in the investigation.

6. Select an hypothesis to be tested.

7. Identify the variable to be manipulated in the investigation.

8. Identify procedures for measuring the variable to be manipulated in the investigation.

9. Identify procedures for manipulating the independent variable in the investigation.

10. Identify the variable to be held constant in the investigation.

11. Identify procedures for holding variables constant in the investigation.

12. Gather data to test the hypothesis.

13. Record data in an appropriate table.

14. Select a suitable scale to graph the data.

15. Plot the data.

16. Interpolate and extrapolate from the graph.

17. Decide whether the data supports the hypothesis.

18. Modify the hypothesis to be consistent with the data collected.

Figure 2. A Linear Model for Planning and Conducting an Investigation. (Tobin and Capie, 1980, p. 595)

## Cognitive Development

According to Piaget's stage dependent theory of cognitive development, most individuals over the mental age of seven, are either at the concrete operational or formal operational stage of cognitive development (Siegler & Richards, 1982). As all subjects to be considered in this study will be fourteen years of age or above, only these stages of cognitive development will be discussed.

At the concrete operational stage, a child's reasoning is orientated towards concrete objects and tasks, the child can only reason about the specific content of the problem. Whereas a child at the formal operational level, does not require concrete objects in order to relate ideas to their cognitive structure.

Piaget has suggested that four factors are necessary for progression through the stages of development: self-regulation, concrete experience, social transmission and maturation of the nervous system. This does not mean that these factors are sufficient for the development of formal thought, "they determine the totality of possibilities and impossibilities at a given stage" (Lawson & Wollman, 1976, p. 413).

Many researchers have noted that small percentages of secondary school students use formal reasoning strategies. Chiappetta (1976) extensively reviewed the studies involving secondary school students in the United States. He concluded that "most adolescents and young adults, do

not appear to have attained the formal operational stage of cognitive development" (Chiappetta, 1976, p. 254). He reported that up to 86% of senior high school students were at the concrete operations level.

Garnett et al. (1985) investigated the reasoning patterns of 1,371 Western Australian secondary school students, aged between thirteen and sixteen. Students' use of proportional reasoning, controlling variables, correlational, probabilistic, and combinatorial reasoning was investigated. They found that in each of the modes tested, the majority of students did not use formal reasoning patterns.

At the formal operational level, the child is capable of separating the variables of a problem, and considering the interaction between them (Levine & Linn, 1977). This was illustrated by Inhelder and Piaget's (1958) pendulum task. During the task children were shown a series of metal balls hung by strings. The weights of the balls and the lengths of the strings differed, giving a large range of different configurations. The task was to determine what factor or factors, affected the period of swing for a pendulum. The length of the string was the only variable which affected the period, this was discovered experimentally by the majority of formal-level children. Almost always the concrete operational child chose weight as the only important factor, or concluded that weight and length were both important. They reached this conclusion,

because unlike their formal operational counterparts, they failed to vary the value of one variable systematically while holding the others constant.

Many studies have been conducted to investigate whether or not the formal activity of controlling variables can be taught (Bredderman, 1973; Lawson, Blake & Nordland, 1975; Sneider, Kurlich, Pulos & Friedman, 1984; Lawson & Wollman, 1976). The training studies reviewed here do not concern training students in all aspects of formal thought, but specifically in the area of variable control. The studies range in their methods but arrive at similar results.

Bredderman (1973) conducted tests on 27 fifth and sixth grade students who had shown inability to control variables in a pretest. The study also examined the relative effectiveness of external reinforcement and cognitive conflict based treatments.

The reinforcement method involved students carrying out a task, in which the results gained reinforced the concept of variable control. With the cognitive conflict model, two or three variables were changed simultaneously, resulting in a misleading relationship between the variables. The child then had to resolve the conflict, by conducting his or her own experiments. Bredderman found that the treatment groups performed slightly higher than the control group on a posttest, but no difference was found on a retention test a month later. He also found

that the two treatment methods used produced little difference in outcomes. It is likely that the young subjects were not sufficiently mature to benefit from the training.

Lawson et al. (1975) used three different tasks to train 33 biology students with a mean age of 15.5 years, to see if the ability to control variables can be taught, and transferred to novel tasks. After training, they found that the students had improved in achievement on the trained task, but no significant differences were found when presented with novel tasks.

Lawson and Wollman (1976), with 32 fifth, and 32 seventh grade students, found that training can assist the development of a formal control of variables strategy, and this strategy once learnt, can be transferred to other specific but novel control of variables tasks. Similar results were obtained by Sneider et al. (1984), who tested the effectiveness of a programme designed to teach children how to plan, conduct, and interpret a controlled experiment. The programme used 275 children aged nine to fifteen from United States schools and non-school youth groups. Posttests illustrated that transfer had occurred, the programme showed that subjects were able to apply the control of variables strategies to a wide range of tasks.

The studies discussed here are diverse, in terms of the modes of instruction and methods of evaluation, it is

therefore difficult to draw precise conclusions. It does appear however, that the ability to control variables is one that can be taught to secondary students, and once learned, the skill can be transferred to other tasks different from those considered during training. In all the studies considered, training involved stressing the need to "change one thing at a time", this is the general premise of variable control. The purpose of this study, is to determine the specific difficulties that adolescents have in planning solutions to problems requiring this skill, and analyze these in light of behaviour exhibited by experts. This analysis will provide an indication of precisely what needs to be taught to promote the acquisition of this valuable skill.

The literature reviewed thus far provides an insight into the mental processes that individuals employ during problem solving activities, such as planning a scientific experiment. Some of the different approaches employed by experts and novices during such a task have been identified.

The review has also described the results of research on the process of planning. In particular the characteristics which distinguish expert planning from that of novices have been discussed. A model of planning specific to experimental design is presented as this provides an initial conceptual framework for analyzing the data collected during the project.

Finally the literature concerning cognitive development was reviewed. The developmental readiness associated with with some of the complex process skills needed in experimental design are discussed. In particular the cognitive skill of controlling variables is described, and studies on the acquisition of this skill are reviewed.

## Methodological Issues

The purpose of this study is to identify the knowledge and strategies used by novices and experts in planning a controlled experiment. A methodology appropriate for this purpose is the thinking aloud approach. This approach, suggested by Larkin and Rainard (1984) aims to provide an understanding of how people approach intellectual tasks. The process involves collecting data on thinking processes, and using these data to model the steps that people take, why they take them, and the errors they make (Larkin & Rainard, 1984).

A simple technique to collect the data is to ask subjects to perform the task, and to think aloud during the task, verbalizing their thoughts as they occur. When such data are collected and transcribed, they are referred to as "verbal protocol data" (Larkin & Rainard, 1984, p. 236).

### The Verbalizaton Process

The process of verbalization externalizes a portion of the information that is currently being processed in STM (Ericsson & Simon, 1980). According to Ericsson and Simon,

the verbalization process puts different demands on STM depending whether or not the information is already represented in verbal code. Verbal information can be accessed from LTM and vocalized automatically without making any further demands on STM. Whereas information such as visual imagery, needs to be encoded into verbal representations, and this does place additional demands on STM, and as a consequence the problem solving process may be slowed down.

Flaherty (1979), gave a series of problems to 100 senior high school students, half of the sample were required to think aloud while solving the problems, and the other half solved them in the usual way. He found no significant difference between the scores attained by the two groups, and concluded that thinking aloud does not affect the problem solving process.

Ericsson and Simon (1980) also argued that thinking aloud does not change the structure of task processes, as verbal reference is made to the structure of STM at any given time, but does not extend the current content of STM. If however, explanations are asked for by the interviewer, current information and relations in STM are interfered with, and this can change the cognitive process. This implies that only neutral probes should be used, this issue is further discussed in the following section.

## Collecting Protocols

In useful "think aloud" protocols the solver talks

steadily, reflecting continuously on what he or she is currently doing, with only a very small amount of neutral input from the interviewer (Larkin & Rainard, 1984). As mentioned above, any comments made by the interviewer can alter the natural thinking processes of the subject.

Larkin and Rainard warned that inexperienced interviewers, particularly teachers, talk too much in the interview, and that comments such as "fine" or "that's good", can inadvertently give subjects the impression that what they are doing is correct. To encourage the subject to keep talking, they have recommended that the interviewer should prepare a list of neutral comments, and only use these during the course of the interview. Larkin and Rainard (1984, p. 250) suggested that comments such as: "Can you say what you're thinking", "That's clear", and "You're talking well" are suitable for this purpose.

Some form of probing is essential to evoke clarification and encourage elaboration (Murphy, 1980). In the think aloud situation, such probes should be left until the end of the interview, and still be as neutral as possible. These questions should not refer to the actual answer arrived at, but the process used on the way, possible questions are: "Can you say more about that?", and "Can you summarize how you got your answer?" (Larkin & Rainard, 1984).

### Modelling Protocols

Simply reading a protocol can give several insights into the the strategies used by the problem solver. However as the protocol can only truly represent a small fraction of the solver's thinking, more is needed to gain a complete understanding of the solver's strategies. The information processing approach, requires categorization of the statements from all collected protocols, and using these to develop a set of coding categories to be used in analyzing each individual protocol.

Each of the problem solving behaviours present in the protocol must be coded by the investigator. Larkin and Rainard recommended that two or more people should be involved in the coding, and that any discrepancies should be debated until consensus is reached.

### Validity Concerns

Internal validity. Guba (1977) cited McCall and Simons'(1969), suggestion that there are four main invalidating factors that may occur during an interview. Distortions resulting from the presence of the researcher, distortions resulting from the researcher's involvement with the subjects, distortions caused by bias on the part of the researcher, and finally distortions arising from the manner in which the data are collected.

The possible interference that may result due to the presence of the interviewer, was discussed above. Theory suggests that provided all comments made by the interviewer

are neutral in nature, very little distortion of natural thinking processes will occur while thinking aloud.

The experimental technique involved interviewing subjects unknown to the interviewer. Also the data collected were purely objective in nature, and it is unlikely that the results were distorted due to involvement with the subjects or through bias.

As mentioned earlier, there can be validity problems arising from incorrect coding of protocol statements. This can be kept to a minimum, if several people are involved in the coding process, and consensus is reached on the coding of each statement.

<u>External validity.</u> Larkin and Rainard (1984) discussed two external validity concerns associated with the information processing approach to research. These problems relate to the typically small sample size involved. They concern the generalizability of the results to a population of interest, and the problem of random sampling.

The behaviours used by the subjects in solving the problem, are only considered if they are observed in a significant number of protocols. Larkin and Rainard argued that this provides strong evidence that the model is applicable to the wider population. They used a binomial probability argument, to show that if the likelihood of

observing a particular rule in any given subject is less than 0.5, the probability of this rule being observed in the majority of individuals in any group purely by chance, is extremely low.

Another common validity problem is that of purely random sampling. Larkin and Rainard accepted this as a genuine concern. Many studies are carried out on intact classes. This raises serious doubts about validity, as any class shares common characteristics that are perhaps unique to that class and their teacher's particular approach to instruction. Techniques should be employed to ensure that the particular population of interest is sampled as randomly as possible.

The review of the literature presented here, forms a theoretical framework to guide the design of the investigation discussed below. It will also provide a basis for the analysis and discussion of results.

# CHAPTER 3

# Method

The design of the study was guided by the research questions in Chapter 1, and the science education literature reviewed in Chapter 2. This chapter considers the choice of technique used to gather the data, as well as describing the development of the instrument used in the study, the selection of subjects, interviewing procedures, data analysis techniques, and the method used to assess scorer reliability.

## Selection of Data Gathering Technique

It was important that the data gathering technique revealed how the subjects worked through the task of designing an experiment, while being as independent as possible of their level of written or oral communication.

The science education literature (Larkin & Rainard, 1984), indicates that the use of think-aloud interviews is the technique which best meets the needs of the study.

## Development of Interview Instrument

Early in Semester II 1989, a pilot study was conducted to gain experience in interviewing, and to gain an insight into some of the difficulties that students have in designing controlled experiments. Two Year 10 and two Year 12 students were used for this purpose. An interview instrument was developed and trialed, and several weaknesses in this initial instrument were identified.

The task chosen for the research, was to ask the subjects to design an experiment to investigate the factors which affect the burning rate of candles. The task was chosen because it was of a novel nature, and not part of the school science curriculum. Because it was novel, the subjects' approaches would not depend, to any great extent, on their subject-matter knowledge.

The interviews were tape recorded and analyzed. The initial interview question was " You are to plan an experiment or series of experiments to investigate which factors affect how quickly candles burn away". It was found that the interview task was too vague and that not all of the students focused on differences between the candles. Rather they tended to discuss external factors which may influence candle burning rate, such as the amount of oxygen. To focus the subject's thinking on the differences between the individual candles, the task was changed to "Candles differ in many ways as you can see. You are to plan an experiment or series of experiments to find out which factors influence how quickly a candle burns away".

The subjects often asked to be reminded of the task while they were working through it. So it was decided to provide each subject with a large card stating the task, which could be referred to during the activity.

The pilot study also revealed that the subjects did

not necessarily describe each step of their plan, such as how the results of their experiment could be used to answer the initial problem. Debriefing questions were constructed to address these problems. The interview instrument including the debriefing questions is presented in Appendix 1.

## Selection of Subjects

Three categories of individuals were used in the study: year 10 science students, year 12 science students and science lecturers who had completed doctoral studies in a science discipline. Six subjects were chosen from each of these categories.

The science students were chosen from a co-educational Catholic school in the Perth metropolitan area. The Year 12 population consisted of all students enrolled in TEE physics, chemistry, biology or human biology. The year 10 population were all students enrolled in science units that specifically lead to TEE science subjects. This controlled for possible aptitudinal differences between the two populations.

Each of the student populations consisted of approximately eighty students. Random stratified sampling was used. This involved dividing each population into three groups according to achievement, and randomly choosing one male and one female from each group.

The sampling of science lecturers could by no means

be done in the random fashion described to select the students. This was due to their comparative short supply. Science lecturers at a Western Australian tertiary institution were used in the study. The only sampling procedure used, was that three lecturers were chosen from each of biological science, and physical science. This was to eliminate the possibility of favouring a particular domain, as this could introduce biases due to approaches that may be specific to a particular science discipline.

## Interviewing Procedures

Each subject was given the task of designing an experiment to examine what factors affect the rate at which a candle burns. Equipment was available for subjects to prompt their planning, but they were not required to actually carry out the experiment, as the study was one of experimental design not implementation. The equipment available at the interviews was a wide range of candles of various shape and colour, matches, a ruler, and pencil and paper. A pocket-sized tape-recorder with an external microphone was used to record the subjects' approach to the problem.

For the student phase of data collection, each student was withdrawn from his or her normal science class for approximately twenty minutes. The student was then taken to a small room in the science area. On the way to the room rapport was developed by casually explaining that the purpose of the interview was to gather information about

the difficulties that students have in designing experiments. It was stressed that the results of the interview had nothing to do with the assessment procedures of the school.

Once in the interview room, with the subject's consent, the tape-recorder was switched on. This was not placed in a prominent position, and a small unobtrusive microphone was used. This was to encourage the student to feel at ease and hopefully forget about its presence. A practice task was then given to accustom the subject to verbalization. The practice task involved the subject thinking aloud while carrying out some simple addition problems. Then the task was explained, and the available apparatus shown to the subject.

As the science lecturers were from various campuses, it was necessary to meet them in their offices and conduct the interview there. The equipment described above was taken to the interview, and it was conducted in the same way as for the students.

During the interview, it was necessary to continually encourage the subject to keep talking. As discussed in the literature review, it was important to ensure that any comments made were neutral in nature and did not affect the subject's thinking processes. Comments such as "that's a good idea", and even "good", were unacceptable as they could have given the subjects an impression about the accuracy of what they were doing. Instead, comments like

"you're explaining your thoughts very well", and "what are you thinking?" were used.

At the end of the interview, the subject was asked to clarify any points of his or her plan which may have been unclear. During this stage, the subject was also asked several debriefing questions to gauge their understanding of a controlled experiment.

## Analysis of Data

Each interview was then transcribed using a dictaphone-type transcriber. This is a play-back device with two foot pedals for forward and reverse. The use of this device removed the need for tedious pausing and rewinding during the transcribing procedure.

Once transcribed, the protocols were segmented into episodes representing single planning operations. Preliminary coding categories were developed by using the model of experimental design proposed by Tobin and Capie (1980). After analyzing several of the transcripts with these initial categories, some categories were removed and others added to develop a coding manual which best matched the planning behaviour demonstrated by the subjects. The actual coding manual is presented in Appendix 2, and the main categories are listed in Figure 3.

| Coding category | Behavior |
|---|---|
| C1 | Identifies variables |
| C2 | States assumptions about variables |
| C3 | Plans measurement procedures |
| C4 | States prediction |
| C5 | Designs an experiment |
| C6 | Plans for control of variables |
| C7 | Plans for repetition of results |
| C8 | Plans for data recording/presentation |
| C9 | Plans for interpretation of data |

Figure 3. Data coding categories.

Once the coding manual was finalized, all of the eighteen protocols were coded. A subject was accredited with a behaviour on the first occasion it was evident, and coding for that behavior was not repeated on subsequent occasions. A sample coded protocol is presented in Appendix 4, to familiarize the reader with the process.

Responses for the debriefing questions were marked according to the marking key provided in Appendix 3. For purposes of analysis and discussion, data gained from the coding and the debriefing procedures were combined and re-organized into separate categories. These were: identification of variables; hypothesizing and predicting; investigation and control of variables; and measurement, data recording and interpretation.

## Scorer Reliability

Two aspects of scorer reliability were investigated, firstly the consistency of the investigator's coding with another science teacher, and secondly the consistency of the investigator's coding from one occasion to the next.

Another science teacher was familiarized with the coding manual and then asked to code six representative protocols from the study. These protocols were concurrently but independently scored by the investigator. The proportion of agreement was calculated for each protocol, as well as the average scorer agreement.

To assess the consistency of the investigator's scoring, the investigator scored three representative protocols on two occasions one week apart.

# CHAPTER 4

# Results

This chapter presents and describes the results of the investigation. The data presented relate to the various operations associated with experimental planning, as well as scorer reliability of the coding procedure. Detailed subject by subject results for the coding and debriefing procedures are provided in appendices 5 and 6 respectively.

## Experimental Planning Operations

### Identification of Variables

Table 1 summarizes the data regarding the subjects' treatment of experimental variables. The table shows the number of subjects from each category who identified variables in a single episode, as well those who recognized variables as they became apparent during their planning. The number of subjects who stated assumptions about the variables is also reported. Results of two debriefing questions are also included, these show the number of subjects who could adequately explain the term "variable" and those who could provide satisfactory examples.

TABLE 1

Subjects' Responses Related to Identification of Variables

| Coding category or research question | Number of Subjects: Science Lecturers n = 6 | Year 12's n = 6 | Year 10's n = 6 |
|---|---|---|---|
| C1a: Identified variables in one episode | 6 | 5 | 3 |
| C1b: Identified variables throughout plan | 0 | 1 | 3 |
| C2: Stated assumptions about variables | 2 | 0 | 0 |
| Q2: Explained the term "variable" | 6 | 2 | 2 |
| Q3: Provided examples of "variables" | 6 | 4 | 4 |

It was found that all of the expert sample identified relevant variables in a single episode and that this was the first task performed. This was in contrast to the Year 10 sample who showed a tendency to identify variables throughout the task. Two of the expert sample stated assumptions about some of the variables which were difficult to measure, for example that all of the candles were made of similar wax, none of the novice sample made such assumptions.

Upon debriefing it was found that only two Year 10's and two Year 12's could define the term "variable" correctly, however four subjects from each group were capable of giving examples of variables from the experiments they had planned. All of the expert sample

provided correct definitions for the term "variable" and gave appropriate examples.

The mean number of variables identified by the Science Lecturers, Year 12's and Year 10's were 5.5, 4.2 and 3.3 respectively. Figure 4 illustrates these results. A Kruskal-Wallis One-way Analysis of Variance by ranks was used to compare the mean number of variables identified by the three groups. Overall group mean scores were significantly different, $H(2) = 6.50$, $p<0.05$.

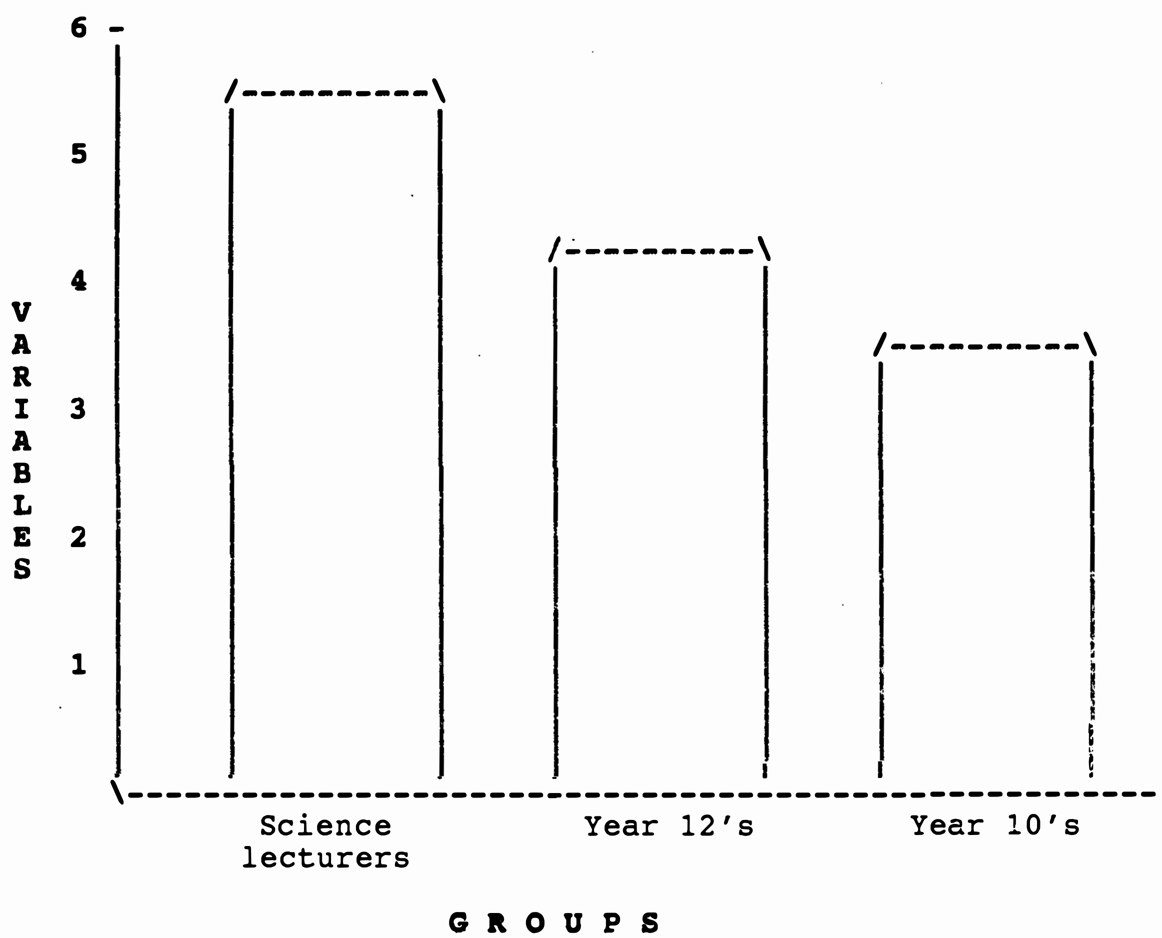


FIGURE 4

Mean number of Variables Identified by each Subject Group

Hypothesizing and Predicting

Hypothesizing and predicting are two similar but subtly different processes involved in experimental design. No coding category was devised for the stating of hypotheses, as none of the subjects did this during their design, so debriefing questions were used to address this problem. Table 2 shows the number of subjects in each group who stated predictions. This table also reports the results of the debriefing questions used to assess the subjects' ability to explain the term "hypothesis", and secondly, their ability to state an hypothesis for their own experiment.

TABLE 2

Subjects' Hypothesizing and Predicting

| Coding category or debriefing question | Number of Subjects | | |
|---|---|---|---|
| | Science Lecturers n = 6 | Year 12's n = 6 | Year 10's n = 6 |
| C4: Stated prediction | 0 | 2 | 3 |
| Q4: Explained the term "hypothesis" | 4 | 2 | 0 |
| Q5: Stated own "hypothesis" | 5 | 1 | 0 |

None of the expert sample stated predictions during their experimental design, compared to two Year 12's and three Year 10's. The predictions generally related to the anticipated outcomes of the experiment. For example, one of the year 10's (student 11) stated: "so the bigger the

candle is umm the longer it will take to burn", this statement was coded as a prediction as it was a specific forecast of a future observation.

The debriefing procedure revealed that four of the experts correctly defined an hypothesis, and five could adequately describe the hypothesis they were testing. This compares with two Year 12's defining the term correctly, and one describing their hypothesis. All of the Year 10 sample were incapable of defining an hypothesis or describing an hypothesis of their own.

## Planning for Investigation and Control of Variables

Vital to any experimental plan, is planning how the identified variables will be treated in the context of the experiment. Table 3 concerns the data related to how the subjects actually planned their investigations and allowed for the control of variables. The results for the debriefing question concerning the subjects' understanding of a controlled experiment are also presented.

TABLE 3

Subjects' Planning of Investigations and Control of Variables

| Coding category or debriefing question | Number of Subjects: Science Lecturers n = 6 | Year 12's n = 6 | Year 10's n = 6 |
|---|---|---|---|
| C5a: Stated experimental design(s) involving one variable | 6 | 4 | 6 |
| C5b: Stated experimental design(s) involving more than one variable | 1 | 2 | 0 |
| C6: Planned controls for most interfering variables | 6 | 2 | 3 |
| C7: Allowed for sample size/repetition | 6 | 1 | 1 |
| Q1: Explained meaning of a controlled experiment | 6 | 3 | 1 |

All of the subjects except two Year 12's, stated an experimental design which involved investigating one variable at a time. For example Student 7 stated:

> if we take a birthday candle and a bigger candle if we umm put them together because they're different umm length and if we burn them umm the small one will burn quicker

One of the science lecturers discussed the possibility of extending the design to test more than one independent variable in the same experiment, while still adequately controlling interfering variables:

> I can combine a number of these variables in the one experiment so I can set up a multi-factorial experiment umm which is more efficient than doing it one variable at a time

All of the science lecturers planned controls for most interfering variables, that is at least two. Whereas only two of the Year 12's and three of the Year 10's demonstrated this behaviour. Characteristic of the subjects who did not control variables was student 10 who stated:

> umm gather a large sample of different candles of different sizes different colours different width and heights different thicknesses of wicks place them in different rooms that have different variables such as wind umm different moisture in the air umm light them at approximately the same time and see how long it takes for them to burn down to what's left

Characteristic of the expert's control of variables was Lecturer 5:

> what I need to do is take groups in which only one thing varies right and test them against each-other ...... so let's say different widths first of all with the same wick and so on and see if the width affects how rapidly it burns then another group taking different wick lengths

It was found that only three Year 12s' and one Year 10 could provide adequate explanations of a controlled experiment. An example of a correct response is provided by Student 2:

> umm a controlled experiment is like where you're only experimenting with one umm like you're keeping all the variables the same except the one you're testing like so you're controlling all the others but like changing which ever one you're testing

Many of the students had poor understandings of controlled experiments, for example Student 4 stated:

> a controlled experiment is done under conditions where there will be the least amount of error and done with the best accuracy you can do and umm spose you could say it was so it would be accurate and safe

All of the expert sample provided appropriate explanations of a controlled experiment, Lecturer 4 provided a particularly succinct definition:

> ah it's an experiment in which all variables are kept constant except the variable of interest

One of the most definite trends revealed by the study related to subjects allowing for repetition of their results for greater experimental accuracy. It was found that all of the expert sample found a need to repeat their experiments, or at least have a large enough sample to improve experimental accuracy. This is illustrated by Lecturer 2: "I suspect this would give a fair bit of variation so we'd probably need to do several trials". Only one subject from each of the student samples made allowance for repetition in their designs.

The mean number of variables each group of subjects planned to treat in their experiments are illustrated in Figure 3. These were 4.7, 3.7 and 2.3 for the Science Lecturers, Year 12's and Year 10's respectively. These results are illustrated in Figure 5.

A Kruskal-Wallis One-way Analysis of Variance by ranks was used to compare the means. Overall group mean scores were not found to be significantly different, $H(2) = 4.69$, which is not significant at the $p<0.05$ level. However a Mann-Whitney test showed that the mean for the science lecturers was significantly different to that of the year 10 students, at the $p<0.05$ level of significance.

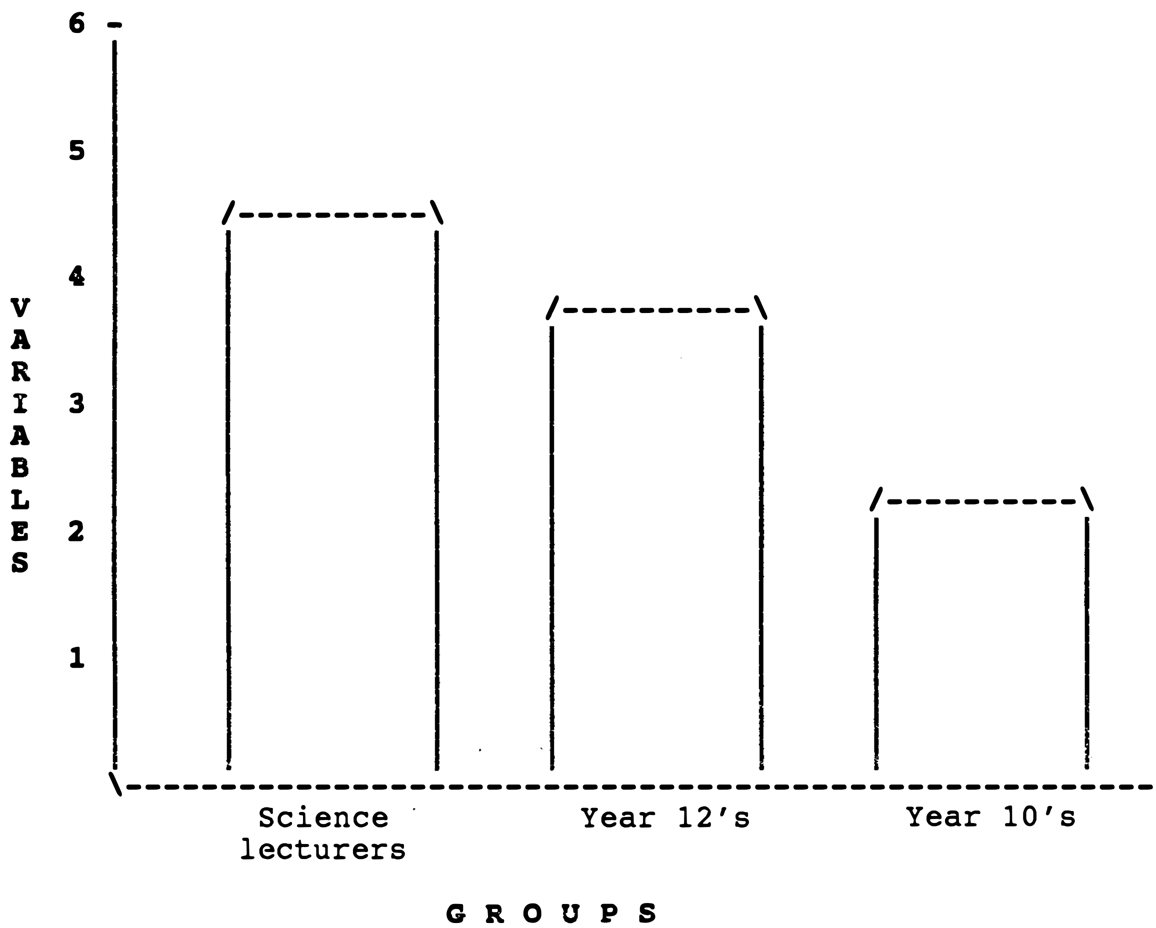


FIGURE 5

Mean Number of Variables Planned to be Tested by Each Subject Group

Planning for Measurement, Data Recording and Interpretation

Table 4 shows the number of subjects in each category who adequately planned for measurement of variables, recording and interpretation of data. The results of two debriefing questions are included, as not all subjects actually made reference to evaluating their hypothesis or measuring their experimental variables.

TABLE 4

Subjects' Planning for Measurement, Data Recording and Interpretation

| Coding category or debriefing question | Number of Subjects: Science Lecturers n = 6 | Year 12's n = 6 | Year 10's n = 6 |
|---|---|---|---|
| C3: Planned measurement procedure(s) for variables | 5 | 6 | 5 |
| C8: Planned data recording/presentation | 1 | 1 | 2 |
| C9: Planned data interpretation | 4 | 3 | 2 |
| Q6: Demonstrated ability to evaluate own hypothesis | 5 | 1 | 0 |
| Q7: Demonstrated ability to plan appropriate measuring technique | 6 | 6 | 5 |

Generally all three groups scored similarly in these categories, except for question 6. One of the expert subjects did not describe his/her measuring technique during the interview, but did so upon debriefing. It was found that the ability to evaluate an hypothesis differed between the groups, with five of the expert sample, one of the Year 12's and no Year 10's scoring in this category.

## Scorer Reliability

The intra-scorer check involved the investigator scoring three representative protocols on two occasions one week apart. It was found that the coding was identical on those two occasions.

The results of the scorer reliability check are presented in Table 5. A colleague was used to test the coding of six representative protocols from the study. The table shows all of the coding categories and all of the debriefing questions for the six subjects. The proportion of agreement is shown separately for each coding category, and for each subject. The figure in the bottom right-hand corner indicates the average inter-scorer agreement for the reliability check.

For the entire inter-scorer check, there were a possible 108 separate codings, there was a discrepancy between the coding of the investigator and that of the colleague on only four of those 108 codings. Overall the inter-scorer reliability of the investigation was found to be 0.97, which indicates that the coding of the data was sufficiently reliable.

TABLE 5

Scoring of Six Protocols by the Investigator and a Science Teacher.

| | Science lecturers 1 | Science lecturers 2 | Year 12'S 1 | Year 12'S 2 | Year 10'S 1 | Year 10'S 2 | Proportion agreed codings |
|---|---|---|---|---|---|---|---|
| Coding categories | | | | | | | |
| C1a | * # | * # | * # | * # | | * # | 1.00 |
| C1b | | | | | * # | | 1.00 |
| C2 | * # | | | | | | 1.00 |
| C3 | * # | * # | * # | * # | * # | * # | 1.00 |
| C4 | | | * # | | | * # | 1.00 |
| C5a | * # | * # | * # | | * # | * # | 1.00 |
| C5b | * # | | | * # | | | 1.00 |
| C6 | * # | * # | * # | | | * # | 1.00 |
| C7 | * # | * # | | | | | 1.00 |
| C8 | | | | | | | 1.00 |
| C9 | * # | * # | * # | | # | | 0.83 |
| Debriefing questions | | | | | | | |
| 1 | * # | * # | | | * # | * # | 1.00 |
| 2 | * # | * # | | * # | * # | # | 0.83 |
| 3 | * # | * # | * # | | * # | | 1.00 |
| 4 | * # | * # | | | | | 1.00 |
| 5 | * # | * # | | * # | | | 1.00 |
| 6 | * # | * # | # | * # | * # | * | 0.67 |
| 7 | * # | * # | * # | * # | * # | * # | 1.00 |
| Proportion agreed codings | 1.00 | 1.00 | 0.94 | 1.00 | 0.94 | 0.89 | 0.96 |

<u>Note</u>
* Denotes scoring by investigator
# Denotes scoring by science teacher

# CHAPTER 5

# Discussion

In this chapter the results of the study are discussed at length, and the difficulties encountered by the novice subjects are considered in light of the literature, and the behaviour of the expert subjects.

## Experimental Planning Operations

### Identification of Variables

There was a tendency for the experts to identify more variables for testing than the novices. The average number of variables identified by the expert sample was 5.5, compared to 3.3 for the Year 10 group and 4.2 for the Year 12's. This suggests that the expert approach was more exhaustive than that of the novices (McGaw & Lawrence, 1984).

It was found that all of the expert subjects identified relevant variables in a single episode and that this was the first task performed. This was in contrast to the novice sample, particularly the Year 10's, who showed a tendency to identify variables throughout the task. Three out of the six Year 10 subjects approached the task in this way.

This result is characteristic of expert-novice differences. As discussed in the literature review,

experts employ a "forward-working" strategy (McGaw & Lawrence, 1984, p. 4). This strategy involves starting with what is known about the problem, and then using this information to work towards a currently unknown end. That is, listing the variables and then designing an experiment based on those variables.

Novices tend to work backwards from the unknown end McGaw & Lawrence 1984). Thus in this case, the novices had their own conceptions about what the outcome of the experiment would be. For example they may have pictured a short candle burning faster than a long candle, and planned an experiment to "prove" this.

The fact that the novices tended to identify variables throughout the planning procedure also supports research by Lawrence et al. (1983). They found that during planning, novices worked in a random fashion and concentrated on immediate tasks, whereas experts worked at a more abstract level, which made their planning more efficient. That is the novices tended to plan for situations as they arose, whereas experts tended to employ an "up-front" planning strategy.

Upon debriefing it was found that only two Year 10's and two Year 12's could provide an accurate definition of the term "variable", however only four subjects from each student group were capable of giving examples of variables in the experiments they had planned. Thus the concept of a

variable was not known to all of the student sample. All of the expert sample provided both an accurate definition of a variable and examples in the context of their experiments.

Two of the six expert subjects made assumptions about some aspect of their experiment, whereas none of the novice sample did this. The assumptions related to variables which were difficult to measure and assumed constant. Generally the assumptions were that all of the candles were made of the same type of wax and had the same density. As only two of the experts made assumptions, it is not a significant trend. However it does demonstrate one of the characteristics of experts discussed in the literature review, that is the completeness and conclusiveness of their solutions (Hackling & Lawrence, 1988).

### Hypothesizing and Predicting

Originally this coding was designed to determine whether or not the subjects stated an hypothesis, however none of the subjects stated hypotheses. It was noticed that several of the novice sample rather than stating an hypothesis in the appropriate way, stated predictions. For this reason the coding category for predictions was created, and the question of stating hypotheses was tackled by the debriefing questions.

An hypothesis is an integrated process skill which involves a general statement of relationship between variables (Funk, Okey, Fiel, Jaus & Sprague, 1979). The

debriefing sessions showed that four of the expert sample held this conception of an hypothesis, whereas none of the Year 10's and two of the Year 12's held the same view. When asked for the hypothesis being tested, no Year 10's and only one Year 12 gave statements sufficiently general to be coded as hypotheses. Whereas five of the expert sample provided general hypotheses.

Three Year 10's and two Year 12's stated a prediction concerning the outcome of their experiment. A prediction differs from an hypothesis in that it is a specific "forecast of what a future observation might be" (Funk et al., 1979, p. 53). It was interesting to note that none of the experts made predictions. Expert planning is characterized by a forward-working approach that leads toward an unknown end. A specific prediction concerning the outcome of an experiment, would imply that experiment needs to be worked back from that point. This is characteristic of novice rather than expert planning. Thus this result clearly illustrates a characteristic difference between expert and novice behaviour identified by (McGaw & Lawrence, 1988).

Planning for Investigation and Control of Variables

Provided that a sufficiently wide range of candles is used, it is quite possible to plan a single experiment which is capable of determining the effect of many different variables on the burning rate of a candle. Of course it would be necessary to have pairs of candles which

only differed in one way in order to control for all of the interfering factors.

The majority of subjects chose instead to plan a series of experiments, each of which investigated the effect of a single variable. All of the expert sample planned their experiments in this way. Although, one suggested that once several variables had been investigated individually, it would be possible to set up an experimental design which allowed for the investigation of many variables at the same time, a "multi-factorial" investigation. Such a design would be valuable for running repetition experiments.

Two of the Year 12 sample planned experiments which investigated several variables in the one experiment. But as can be seen from Appendix 5, these two subjects failed to plan any controls. That is, they wanted to use a wide range of candles, light them all at the same time and see which burnt the fastest. If a red, short, ornate candle burnt quickest, presumably to them, these would be the characteristics of fast burning candles.

As discussed in the literature review, the ability to control variables is one that is usually attained after a child has reached the formal operational level of reasoning (Inhelder & Piaget, 1958). Many studies have shown that the majority of school aged children are incapable of applying formal reasoning patterns (Inhelder & Piaget, 1958; Chiappetta, 1976; Garnett et. al., 1985). Therefore it is

not surprising that so few of the student sample adequately controlled variables during their experimental design. All of the six experts planned controls for interfering variables in their designs, this is compared with three of the Year 10 sample and two of the Year 12 sample.

During the debriefing, only three Year 10's and one Year 12 provided adequate explanations of a controlled experiment. Many of the students believed that a controlled experiment was one that was accurate and safe. Others simply said that they had no idea what was meant by the term.

The results do tend to support the literature concerning the proportion of adolescents who have attained formal reasoning. Although this study did not involve training the participants, the literature suggests that the ability to control variables is one that can be taught. Though students are introduced to the concept of a controlled experiment in Year 8, and this concept is reinforced throughout the high school years, there appears to be a definite need to improve instruction in this area.

There was little or no relationship between the novice subjects who had a good definition of a controlled experiment, and those who planned a good controlled experiment during the task, this is shown Appendices 5 and 6. The students who performed well on the planning task had a strong grasp of the concept of a controlled

experiment, as they demonstrated this by their planning, however they had little or no working definition for the concept. This could suggest that these students had not been taught a formal definition of a controlled experiment.

All of the experts made some reference to the need for repetition in their experimental designs in order to make their results reliable. Only one of the Year 10's and one of the Year 12's, discussed the need for repetition. This result is extremely significant, as previous models for planning and conducting an experiment (Tobin & Capie, 1980), made no reference to this stage of an experimental design.

Generally in the classroom situation, time constraints render the repetition of experiments impossible. However in the scientific community, repetition is crucial to the validity and reliability of results. Thus this result may not be indicative of expert/novice differences, as much as it illustrates a key difference in how science experiments are treated in the school setting, compared to how they are carried out in the scientific world.

There appears to be a need to include experiments which involve repetition in the school science curriculum. If school science is giving students the impression that a single experiment is sufficient to support or disprove an hypothesis, there is obviously something very wrong with the way experimenting is currently taught. One technique which can be used by science teachers to increase the sample size of experiments, is to pool data collected by

different groups. Provided each group uses the same experimental method, this is a suitable technique for increasing the number of trials for an experiment.

## Planning for Measurement, Data Recording and Interpretation

All but one novice and one expert planned appropriate measuring procedures for their experiments. The idea of measuring the burning rate of the candle was so implicit in the task, it is not surprising that the vast majority of subjects performed so well in this category. It was presumably this which led to one expert neglecting to state his/her intended measuring technique. The seventh debriefing question was used to assess the measuring techniques used by all of the subjects whether or not they were described during the interview. Responses for this question supported the results of the coding procedure.

There were some differences in the techniques used to compare the burning rates of the candles. The experts tended to plan quantitative procedures which would result in a mathematical expression, for example a rate in centimetres per hour. The novices were more inclined to plan a qualitative experiment which involved lighting two candles simultaneously and simply noting which one burnt out first, this of course is quite a suitable method to employ, but is not so useful when comparing results between different experiments.

Only one of the expert sample, two Year 10's and one

Year 12 made direct reference to the technique they planned to use for recording or presenting their data. Nearly all of the subjects made some reference to recording data, but didn't mention setting up a table or drawing a graph. This category was included for the sake of completeness, and it was hoped that there would be some differences in the way the expert and the novice groups chose to record their data. The task did not require the subjects to state their data recording procedure, and unfortunately it was not included as a debriefing question. As a consequence there is very little to comment upon for this coding category. However it does appear that data recording is generally planned at the time of taking measurements, and not considered during up-front planning.

To be accredited with planning for data interpretation, the subject had to state how the data, once collected, could be used to answer the initial problem. Four of the expert sample, two of the Year 10's and three of the Year 12's planned for data interpretation. How the data were to be interpreted was fairly implicit in the task, and this could explain why such low numbers of subjects made reference to it.

This outcome was anticipated after the pilot study, so a debriefing question was developed to assess the planned data interpretation. This question was "At the end of your experiment, how will you know if your hypothesis is true or false?", four of the Year 10 sample and four of the Year 12

sample adequately answered this question. All of the expert group responded correctly to this item. Even if the subjects' conception of hypothesis was inadequate, he/she could still score on this item.

Generally the expert subjects tended to plan statistical methods of analysis. This involved generating averages by testing a group of candles of the same type, and then statistically comparing that average to the average burning rate of other types of candles. In contrast the novices tended to simply burn two candles simultaneously, and record which burnt fastest.

It was mentioned in the literature review that experts tend to apply their knowledge, whereas novices don't always apply what they know (Chase & Ericsson, 1981), and that problem solving by experts is generally more exhaustive than that by novices (Hackling & Lawrence, 1988). The techniques the experts planned to use to interpret data, in their simplest form, were not beyond the prior knowledge of the novices. However even those novices who chose to carry out repeat experiments, did not discuss comparing average burning rates. So although there were no differences between the groups in whether or not data interpretation was planned, there were characteristic expert/novice differences in the techniques chosen.

Under current patterns of instruction, students are not provided with sufficient opportunity to plan experiments of their own. The vast majority of science

experiments in the school years, and in many cases at the tertiary level, are based on a prepared list of instructions. These instructions are either generated by the teacher or provided in a laboratory manual, and do little to promote a proper understanding of scientific inquiry.

In view of this current practice, the results of this study are not surprising. Students are not given sufficient practise in experimental design, and as a consequence have low attainment of the process skills associated with planning a controlled experiment.

# CHAPTER 6

## Summary and Conclusions

This chapter summarizes the main findings of the study, and identifies implications for instruction and further research.

### Summary of Findings

The various approaches to experimental design employed by the three categories of subjects used in the study have been presented in Chapter 4, and these are discussed in Chapter 5. These data are summarized here in relation to the research questions of the project.

Generally it was found that the expert subjects had a better understanding of experimental variables than the novices. The experts tended to identify more variables than the novices, and they did this in a single episode at the beginning of their plan, whereas the novices tended to identify variables throughout their plan.

It was found that the novice subjects were inclined to state predictions concerning the outcome of their planned experiment, and generally experienced difficulty formulating an hypothesis of their own. None of the science lecturers stated predictions during their plan, and they were far more capable of formulating hypotheses when asked.

While planning experiments, it was found that all of the experts planned controls for interfering variables, whereas only half of the novice sample did this. It was found that the novice subjects experienced considerable difficulty in explaining the meaning of a controlled experiment.

One of the more definite outcomes of the research was the tendency for the experts to plan for repeated trials so as to increase the size of the sample. This step in experimental planning was not part of the planning model proposed by Tobin and Capie (1980), but all six of the science lecturers interviewed, discussed the need for repeat experiments.

The study revealed few expert-novice differences in planning for measurement, recording and interpreting data. Only one of the novice subjects could adequately evaluate an hypothesis, compared to five out of six for the experts. However this was not surprising in view of the two groups' comparative understanding of hypothesizing discussed above.

<u>Limitations of the Study</u>

There were several limitations to the study. Although a novel task was used to identify the planning characteristics of the subjects, few tasks can be completely independent of the subjects' prior content knowledge. For example it is possible that the experts, due to their age and general life experience, had a better

knowledge of the behaviour of candles under various conditions.

The sample size chosen was very small, as only eighteen subjects were used in the study. Also all of the students were chosen from the same school and all the science lecturers were chosen from the same tertiary institution. Therefore the study has limited generalizability to the general population of school science students and tertiary science lecturers.

## Implications for Instruction

Currently the majority of school science experiments are based on a "recipe' style laboratory manual, which takes students through their "experiment" step by step. This approach does little to foster an understanding of scientific inquiry, as the students are not involved in planning the design of the experiments. Students need to be provided with experiences where they can carry out entire experiments themselves, right from the planning stage through to summarizing data and making conclusions. This way, students will gain a more accurate understanding of the nature of science.

Because of this "cook-book" approach to science experiments, students receive little opportunity to practise the process skills associated with experimental design. Students need to be exposed to more investigative style experiments if they are to develop these important process skills.

It seems students receive little explicit instruction on the nature of variables, and types of variables. Science instruction needs to stress the importance of variables in scientific experiments. Both in terms of their identification and their manipulation. Many of the students identified very few of the relevant variables in the experimental task given to them. Their performance was probably limited by their knowledge of the phenomenon of variables. Students need more practice in analyzing experimental situations, to identify relevant variables which may influence those situations, and thus the results of any subsequent experiments.

With a better understanding of the relationship between variables in an experiment, students will be better equipped to identify any influences that the various variables may have on one another, and hypothesize about any likely relationship between the variables. They would also become aware of the necessity for isolating variables during investigations, and thus be more capable of planning adequately controlled experiments.

Only two of the twelve student subjects interviewed, planned for repetition of results. This highlights a weakness in the way experiments are treated in the school setting. Science teachers need to make students aware of the need to repeat experiments for the sake of experimental accuracy.

Although time constraints place restrictions on the repetition of experiments, it is possible to increase the size of an experimental sample without necessarily allocating extra time. This can be achieved by pooling data from the various groups in the class and calculating means. This approach is valid only if all groups are applying the same experimental techniques, if students are developing their own techniques they would need to incorporate repetition into their plans.

## Implications for Further Research

Further research could be directed at some of the limitations of the current research outlined above. The testing methodology could be modified to pencil and paper tests to allow for larger samples. The results of such a study would check the generalizability of the results of this project. The range of sample groups could be extended to include teachers and student teachers. Results from such a study would have implications for teacher education.

This project did not investigate "on the job" planning, but rather "up front" planning. Future research could be directed toward the adequacy of planning when subjects are asked to plan and carry out an investigation.

## REFERENCES


Anderson, J.R. (1981). Cognitive skills and their acquisition. New Jersey: Lawrence Erlbaum Associates.

Bereiter, C., and Scardamalia, M. (1986). Educational relevance of the study of expertise. Interchange, 17, 2, 10-19.

Bredderman, T.A. (1973). The effects of training on the development of the ability to control variables. Journal of Research in Science Teaching, 10, 3, 189-200.

Chase, G., and Ericsson, K. (1981). Skilled memory. In J.R. Anderson (ed), Cognitive Skills and their Acquisition. New Jersey: Larwence Erlbaum, 141-180.

Chiappetta, E.L. (1976). A review of Piagetian studies relevant to science instruction at the secondary and college level. Science Education, 60, 2, 253-261.

Cohen, G. (1983). The Psychology of Cognition. New York: Academic Press.

Education Department of Western Australia, Curriculum Branch (1981). Syllabus for Lower Secondary Science.

Ericsson, K.A., and Simon, H.A. (1980). Verbal reports as data. Psychological Review, 87, 3, 215-251.

Estes, W.K. (1982). Learning, memory, and intelligence. In R.J. Sternberg (ed), Handbook of Human Intelligence. London: Cambridge University Press, 170-224.

Flaherty, E.G. (1979). The thinking aloud technique and problem solving ability. The Journal of Educational Research, 97, 5, 223-225.

Funk, H.J., Okey, J.R., Fiel, R.L., Jaus, H.H., and Sprague, C.S. (1979). Learning Science Process Skills. Iowa: Kendall/Hunt Publishing Company.

Garnett, P.J., Tobin, K., and Swingler, D.G. (1985). Reasoning abilities of secondary school students aged 13-16 and implications for the teaching of science. European Journal of Science Education, 7, 4, 387-397.

Guba, E.G. (1977). Toward a Methodology of Naturalistic Enquiry, CSE Monograph. Los Angeles: University of California.

Hackling, M.W., and Lawrence, J.A. (1988). Expert and novice solutions of genetic pedigree problems. Journal of Research in Science Teaching, 25, 7, 531-546.

Hayes-Roth, B., and Hayes-Roth, F. (1979). A cognitive model of planning. In R.J. Sternberg, and D.K. Detterman (eds), Human Intelligence: Perspectives on its Theory and Measurement. Yale University Press.

Inhelder, B., and Piaget, J. (1958). The Growth of Logical Thinking from Chidhood to Adolescence. New York: Basic Books.

Larkin, J.H., and Rainard, B. (1984). A Research methodology for studying how people think. Journal of Research in Science Teaching, 21, 3, 235-254.

Lawrence, J.A., Dodds, A.E., and Volet, S.E. (1983). An afternoon off: A comparative study of adults and adolescents' planning activities. Conference Paper Presented at the Annual Conference of the Australian Association for Research in Education. Canberra.

Lawson, A.E., and Wollman, W.T. (1976). Encouraging the transition from concrete to formal cognitive functioning - an experiment. Journal of Research in Science Teaching, 13, 5, 413-430.

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.

Levine, D.I., and Linn, M.C. (1977). Scientific reasoning ability in adolescence: Theoretical viewpoints and educational implications. Journal of Research in Science Teaching, 14, 4, 371-384.

Lindsay, P.H., and Norman, D.A. (1977). Human Information Processing: An Introduction to Psychology (2nd ed). New York: Academic Press.

McGaw, B., and Lawrence, J. (1984). Developing expertise through higher education. HERDSA News, 6, 3, 3-7.

Murphy, J.T. (1980). Getting the Facts. U.S.A.: Scott, Foresman and Company.

Novak, J.D. (1980). Learning theory applied to the biology classroom. American Biology Teacher, 42, 5, 280-285.

Shiffrin, R. M., and Dumais, S.T. (1981). The development of automatism. In J.R. Anderson (ed), Cognitive Skills and their Acquisition. New Jersey: Lawrence Erlbaum Associates.

Siegler, R.S., and Richards, D.D. (1982). The development of intelligence. In R.J. Sternberg (ed), Handbook of Human Intelligence. London: Cambridge University Press.

Sneider, C., Kurlich, K., Pulos, S., and Friedman, A. (1984). Learning to control variables with model rockets: A neo- Piagetian study of learning in field settings. Science Education, 68, 4, 463-484.

Sternberg, R.J., and Detterman, D.K. (eds) (1979). Human Intelligence: Perspectives on its Theorv and Measurement. Yale University Press.

Sternberg, R.J. (ed) (1982). Handbook of Human Intelligence. London: Cambridge University Press.

Stewart, J.H., and Atkin, J.A. (1982). Information processing psychology: A promising paradigm for research in science teaching. Journal of Research in Science Teaching, 19, 4, 321-332.

Stillings, N.A., Feinstein, M.H., Garfield, J.L., Rissland, E.L., Rosenbaum, D.A., Weisler, S.E., and Baker-Ward, L. (1987). Cognitive Science an Introduction. Massachusetts: M.I.T. Press.

Tobin, K.G., and Capie, W. (1980). Teaching process skills in the middle school. School Science and Mathematics, 80, 7, 590-600.

Wingfield, A. (1972). Decay of information in short term memory. Science, 176, 690-692.

# APPENDIX 1

## Interview Schedule

### Introduction

The purpose of this activity is to identify the sorts of difficulties that students have in planning an experiment. The results from this exercise will not be given to the school and have nothing to do with assessment.

### Practise Problem

1. I'm going to ask you to plan an experiment, thinking aloud while you are doing it. I'll show you what I mean by thinking aloud with a simple addition problem...............

2. Here, you try a few.................................... - I'll just turn the tape on.

### Task

1. Candles differ in many ways as you can see. You are to plan an experiment or series of experiments to find out which factors influence how quickly a candle burns away.

2. I have a few candles here, but you could have as many as you liked for the experiment.

3. While you are planning the experiment, I would like you to think aloud, just as you did with those sums.

4. Do you understand exactly what I want you to do?

5. Remember that I want you to continue talking as you work through the exercise, and that you can make notes if you wish.

6. I would like you to read the task out loud, then keep talking as you work through it.

[MAKE NOTES ON POSSIBLE POINTS FOR CLARIFICATION]

-----------------------------------------------

-----------------------------------------------

-----------------------------------------------

-----------------------------------------------

-----------------------------------------------

7. Are you happy with that?

8. There's nothing else you need to plan?

<u>Debriefing</u>

1. Now could you explain for me in your own words, what a controlled experiment is like?

2. Could you tell me what a variable is in an experiment?

3. What were some of the variables in your experiment?

4. Could you tell me what a hypothesis is?

5. Could you tell me the hypothesis you were testing in your experiment?

6. At the end of your experiment how will you know if your hypothesis is true or false?

7. You said that you would measure the rate of burning of the candle. Can you tell me how you would do that?

## APPENDIX 2

### Coding Manual

| CODING | BEHAVIOUR | CRITERIA |
|---|---|---|
| C1a | Identifies variables in one episode. | Clearly identifies some factors which may affect the burning rate of a candle. e.g. width length, shape, wick length. |
| C1b | Identifies variables throughout plan. | |
| C2 | States assumptions about variables. | States an assumption about a variable that may be difficult to measure. e.g. assumes constant density |
| C3 | Plans measurement procedures for variables. | Plans an appropriate procedure for measuring variables. e.g. plans a technique for measuring the rate of burning of the candles |
| C4 | States prediction. | States some specific prediction concerning the outcome of the experiment. e.g. I think that the patterned candles will burn faster than the plain ones |

| CODING | BEHAVIOUR | EXPLANATION |
| --- | --- | --- |
| C5a | Experimental design 1 independent variable | States an experimental design(s) that includes an independent variable that is to be manipulated and a dependent variable that is to be measured |
| C5b | Experimental design more than 1 independent variable | |
| C6 | Plans controls for most interfering variables. | At least two interfering variables must be controlled. e.g. I'd get two candles that have the same length, wick length and colour, but had different shapes |
| C7 | Sample size/ repetition. | Plans for repetition to make the experiment more valid. e.g. I'd have at least three of each type of candle. |
| C8 | Data recording/ presentation. | States a technique for recording or presenting the data. e.g. plans a table or a graph. |
| C9 | Data interpretation. | Explains how the data collected will be used to solve the initial problem. e.g. by comparing results. |

| CODING | BEHAVIOUR | EXPLANATION |
| --- | --- | --- |
| C10a | Number of independent variables. | The number of variables which are treated as independent during the design. |
| C10b | Number of variables mentioned. | All of the variables to which reference was made. |

## APPENDIX 3

### Debriefing Question Marking Key

The following marking key sets out the criteria which were used to score the debriefing questions.

| QUESTION | CRITERIA FOR SCORING |
|---|---|
| 1. Explain for me in your own words what a controlled experiment is like? | Must state that one variable is changed while all others are held constant. |
| 2. Could you tell me what a variable is in an experiment? | States that a variable is a feature of an experiment that can change, or be changed. |
| 3. What were some of the variables in your experiment? | Examples such as height, width, shape, colour, wick length or ambient conditions need to be stated. |
| 4. Could you tell me what an hypothesis is? | Must imply that an hypothesis is a general statement of the supposed relationship between variables. It is not a specific prediction. |
| 5. Could you tell me the hypothesis you were testing? | Must state a general relationship between the variables of the experiment e.g. that the length of a candle affects the rate at which it burns. |

| | |
|---|---|
| 6. At the end of your experiment how will you know if your hypothesis is true or false? | Must state a technique for evaluating their stated hypothesis. e.g. by analyzing the experimental data. |
| 7. You said that you would measure the burning rate of the candle. Can you tell me how you would do this? | Must state an appropriate technique for measuring, or comparing the burning rates of candles. e.g. By burning each candle for one hour, and measuring the length burnt away. |

# APPENDIX 4

## Sample Protocol

All of the eighteen protocols were segmented into episodes, then coded according to the coding manual provided in the appendix. The sample protocol provided here is representative of the expert sample, and is included to provide an example of the raw data collected during the investigation, and familiarize the reader with the coding procedure.

A subject was accredited with a certain behaviour on the first occasion it was evident, and coding for this behaviour is not repeated on subsequent occasions throughout any particular protocol. The codings are shown in the right margin adjacent to the episode in which it was observed.

The protocol provided is that for a typical expert subject; Science Lecturer 4. An expert protocol was chosen to illustrate as many of the coding categories as possible.

This particular subject demonstrates several of the characteristic expert tendencies. For example all of the variables to be investigated were identified at the beginning of the plan (C1a). Also the subject isolated all variables except the one under consideration (C6), that is the subject controlled for interfering variables.

# Protocol for Science Lecturer 4

Umm well I'd start by making a list of umm sort of possible factors that might influence how quickly candles burn / and I'd probably jot down things like ah the environmental factors / perhaps the temperature of the environment of which the ah ah temperature of the environment in which the candles are burning / umm whether the air is still or windy so umm say air movement / umm the width of the candle / perhaps umm the shape of the candle whether it has ah a smooth surface or umm or a sculptured surface / the what else could influence the position of the candle in terms of what umm what wax what material has been used in the candle / ah the nature of the wick umm the wicks are made of different materials that might affect the rate at which the candle is burning / so probably I may if I spend some time on it maybe I could think of more things that might affect the rate at which candles burn / and ah make a list of these / and then I'd attempt to isolate each of these possible variables in an experimental design / so for example if I thought that

C1a: Identifies variables in a single episode

C6: Subject planned controls for most interfering variables

ambient temperature umm might affect the rate at which a candle burnt umm other words I guess that's the temperature of the combustible material in the candle / then I would select two candles which are identical in every respect in terms of their colour their surface their mass to length ratio / umm and I would put those two candles which were identical in every respect into two chambers which were identical in every respect except for the fact that one was a different temperature than the other / I'd light them / measure the rate at which the candles burnt in in terms of number of centimetres per hour / depending on the size of the candle / and ah then would then compare those and see / I'd repeat that experiment a number of times ah sufficiently often to give statistically valid results / and then compare the means of the rates at which they had burnt in terms of the length of the candle consumed per unit time / in response to the actual problem of course I'd check all of those variables by isolating them

C5a: States an experimental design involving one variable

C3: Plans procedures for measuring dependent variables

C7: Plans for repetition of experiment

C9: Plans how the data is to be interpreted

## APPENDIX 5

### Subject by Subject Analysis of Coding Procedure

CODING CATEGORIES

| | C1a | C1b | C2 | C3 | C4 | C5a | C5b | C6 | C7 | C8 | C9 | C10a | C10b |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SCIENCE LECTURERS | | | | | | | | | | | | | |
| #1 | * | | * | * | | * | | * | | * | * | 4 | 4 |
| #2 | * | | | * | | * | | * | * | | | 4 | 5 |
| #3 | * | | * | * | | * | * | * | * | | * | 3 | 6 |
| #4 | * | | | * | | * | | * | * | | * | 7 | 7 |
| #5 | * | | | | | * | | * | * | | * | 4 | 5 |
| #6 | * | | | * | | * | | * | * | | | 6 | 6 |
| YEAR 10'S | | | | | | | | | | | | | |
| #1 | | * | | * | * | * | | | * | * | * | 1 | 3 |
| #2 | | * | | * | | * | | | | | | 4 | 4 |
| #3 | * | | | | * | * | | | | | | | |
| #4 | * | | | * | | * | | * | | * | | 4 | 5 |
| #5 | * | | | * | * | * | | * | | | | 3 | 3 |
| #6 | | * | | * | | * | | * | | | * | 1 | 3 |
| YEAR 12'S | | | | | | | | | | | | | |
| #1 | | * | | * | | * | | | | * | * | 4 | 4 |
| #2 | * | | | * | * | * | | * | | | * | 4 | 4 |
| #3 | * | | | * | | * | | | * | | * | 1 | 2 |
| #4 | * | | | * | | * | | * | | | | 2 | 3 |
| #5 | * | | | * | * | | * | | | | | 5 | 6 |
| #6 | * | | | * | | | * | | | | | 6 | 6 |

<u>Note:</u> * Denotes a particular behaviour coded for a particular subject.

APPENDIX 6

Subject by Subject Analysis of Debriefing Procedure

| | QUESTIONS | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| SCIENCE LECTURERS | | | | | | | |
| #1 | * | * | * | | | * | * |
| #2 | * | * | * | * | * | * | * |
| #3 | * | * | * | * | * | * | * |
| #4 | * | * | * | * | * | * | * |
| #5 | * | * | * | | * | * | * |
| #6 | * | * | * | * | * | * | * |
| YEAR 10's | | | | | | | |
| #1 | * | | * | | | * | * |
| #2 | * | * | * | | | * | * |
| #3 | | | | | | | |
| #4 | | * | * | | | * | * |
| #5 | * | | | | | * | * |
| #6 | | | * | | | | * |
| YEAR 12's | | | | | | | |
| #1 | | * | * | | | | * |
| #2 | | | * | | | | * |
| #3 | | | | | | * | * |
| #4 | | | * | * | | * | * |
| #5 | * | | | * | | * | * |
| #6 | | * | * | | * | * | * |

<u>Note:</u> * Refers to a particular subject answering a particular debriefing question correctly.