# THE OBSERVATION COMPETENCE OF GRADE SIX SCIENCE STUDENTS



CARL NORMAN SHEPPARD, B.Ed.

# THE OBSERVATION COMPETENCE OF GRADE SIX SCIENCE STUDENTS

By

© Carl Norman Sheppard, B.Ed.

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of Master of Education

Faculty of Education

Memorial University of Newfoundland

April 1991

St. John's Newfoundland

ISBN 0-315-68237-X

## Abstract

The primary goal of this study was to examine the ability of grade six students with the science process skill and critical thinking skill of observation. Twenty-four students were interviewed as they worked through a series of science activities that required them to make and report observations. Their reports were analyzed using the various conditions of observation competence in a model by Norris (1984). This model lists various conditions which facilitate good observation. Factors such as the observer being alert, having theoretical understanding, and using precise methods are the types of conditions that are included.

The typical grade six student was found to be considerably lacking in observation ability when probed with non-leading questions. Among the weaknesses, there was a general lack of alertness, theoretical understanding, and poor competence in reporting observations in a record. However, in response to leading questions, there was a much more satisfactory level of competence, except for the area of theoretical understanding which showed no leading probe effect.

Students' reports were used to produce qualitative descriptions of the typical student, as well as of three individuals who represent an average observer, an above-average observer and a below-average observer. These

individual descriptions detail specifically the level of proficiency each student possesses with each category of observation competence, and provides illustrations of how the students displayed this competence in their responses to questions and in their behaviours while conducting the activities.

## Acknowledgement

I gratefully acknowledge the advice and encouragement of my supervisor, Dr. Stephen P. Norris. Throughout the entire procedure of reading early drafts, providing consultation, and making suggestions, his personal commitment and high ideals provided the wherewithal for this, the final product. My sincere thanks and appreciation are offered.

I thank also, Dr. Glenn Clark, who provided many helpful comments during the preliminary stages of this study.

## Dedication

This thesis is dedicated to my wife Jackie and my three children Aldonna, Jamie, and Randy who unwittingly gave so much that I might apply myself to the demands of a graduate program. The missed vacations, and the time I spent away from home are among the sacrifices they made, and for these there can be no adequate compensation.

TABLE OF CONTENTS

Page

## LIST OF TABLES

## LIST OF FIGURES

# CHAPTER 1: OVERVIEW OF STUDY

## The Question To Be Studied

The central purpose of this study is to develop and use an interview approach for conducting a qualitative analysis of the observation competence of a sample of grade six students. These students have experienced nearly seven years of instruction with a program that, it is claimed, promotes the science processes. If the program is effective, they should have attained some degree of competence in making and reporting observations.

The primary aim is to portray the observation competence of the typical grade six student, that is, to determine the competencies such a person possesses or lacks that can influence the accuracy of observations made. For example, are students able to disregard preconceived notions, or at least not let them interfere, when they observe something incompatible with those notions? Are students aware of how various factors, such as access to the thing observed, can affect the accuracy of their observations? Do students tend to make their observations carefully and with the required precision?

In the study, students are observed as they make observations in a series of science activities, and are probed with questions intended to elicit information that can be interpreted in terms of its relevance to observation competence. Through this approach, a picture is drawn of the typical grade six student's ability to make accurate observations. This portrayal could be used as a norm against which to evaluate other students and as a guide in setting goals for observation competence at the primary and elementary science levels.

The secondary aim is to develop an individual interviewing method for appraising the observation ability of science students. Most of the previously used research and evaluation instruments are of the paper-and-pencil variety, and are constructed on the basis of very narrow definitions of the science processes. For example, Hungerford and Miles (1969) designed a test of observation and comparison skills on the basis of the following definition: "The behavior of scientific observation is usually described as the ability to make accurate observations with the subsequent communication of these same observations" (p. 62). However, this definition is not helpful for the type of evaluation in this study because it does not specify just what competencies would assist a person in making accurate observations or in subsequently reporting them. An observation test constructed on such a definition may determine in what percentage of items a person observed successfully, but it will not indicate the qualities the student possessed or lacked that determined performance on the test. The intent of this study is to demonstrate a method that compensates for this weakness in traditional tests.

The study will draw upon a model of observation competence proposed by Norris (1984). The model covers a broad range of proficiencies, categorized into a number of subdivisions, that define a competent observer. Possession of the competencies facilitates a person's ability to make accurate observations. This model is used to determine the types of information sought and, therefore, the types of probing questions to be asked.

## Background to the Study

During the past twenty-five years, promotion of the processes of science has become a major aim of elementary school science programs. This aim was popularized largely through the efforts of Science--A Process Approach (AAAS, 1967-68), an elementary science program sponsored jointly by the National Science Foundation and Xerox Corporation. This promotion of science processes is now a major goal of most science programs. However, the development of instruments suitable for measuring students' understanding of, or facility with, the science processes has not kept pace with the development of curricula. Even though students are exposed to programs promoting the science processes during their primary and elementary school years, teachers have no dependable method for determining to what extent the aim is being met. Furthermore, teachers have no way of knowing how competent students can be expected to be.

The model of observation competence that Norris (1984) has proposed has potential for clarifying the process goals and guiding process evaluations of science curricula. Norris conceives of observation competence as consisting of three broad proficiencies. A competent observer is proficient at: (a) making observations well, (b) reporting observations well, and (c) correctly assessing the believability of reports of observations. In addition, he contends that there is more to making observations well than perceiving carefully, precisely, and thoroughly, and more to reporting observations well than reporting them accurately. In support of this claim Norris presents sets of conditions conducive to making and reporting observations well and a set of principles for assessing the believability of reports of observations. These conditions and principles are tabulated and described at length in the chapter on theoretical framework and methodology.

Finally, Norris claims that meeting these conditions and following these principles facilitate accurate observations. In this study, evidence is presented on the extent to which the sample group does possess these competencies.

## Motivation for the Study

Various processes such as classifying, inferring, controlling variables, and observing may be used as a scientific investigation is carried out. These and other processes are considered worthy of being taught, not only because of the significance they have for science, but also because of their presumed transferability to other areas of the curriculum and to life in general (Gagné, 1963). For example, Holland, Holyoak, Nisbett, and Thagard (1986), in a comprehensive account of induction, have discussed how inference abilities have been found to be transferable. A particular example they discuss is how training in statistics does have definite bearing on how individuals solve problems in other areas.

Recent work by Ennis (1989, 1990), Perkins and Salomon (1988, 1989), and Sternberg (1987) suggests that there are instances where skills can be transferred to other contexts, although they acknowledge the importance of content-specific knowledge on which to exercise those skills.

Opposition to this view that thinking skills and science processes are generalizable can be found. Schoenfeld and Hermann (1982) and Hirsch (1987) argue that content-specific knowledge, not general problem solving heuristics, is the key to success in dealing with situations. Hirsch argues that we should ignore general skills, and instead equip youngsters with the varied basic knowledge that makes one culturally literate.

There are, however, indications of common ground. For example, McPeck (1990), who has been one of the strongest opponents of general thinking skills, has now softened his position and conceded that there are some very limited general thinking skills. Brandt (1990) reports on a new program called "Connections" that is being developed by David Perkins. It confronts the need for conceptual understanding of subject matter on one hand and the need for general thinking skills on the other. It's a program designed to help teachers integrate the teaching of particular thinking strategies with their subject matter instruction. There is much to be resolved in this area, however it is becoming more accepted that knowledge and skills learned in one context can be applied to other situations.

Observation is viewed as the most fundamental of the science processes (AAAS, 1967-68). The success or usefulness of any scientific investigation is determined by the accuracy of the observations that are made. The most carefully planned experiment is useless if it depends on observations that are inaccurate or inadequate. If, by promoting the process of observing in a science curriculum, we can also make a person a better observer in all life situations, then the emphasis on the use of the process learning approach in science is justified.

The science processes have not, however, been studied in depth. Much of what is commonly accepted as true, such as the transferability of the skills, has little empirical support. In addition, we do not have a clear conception of the degree of competence that students can acquire with a skill such as observing. We do believe that students should leave school as competent observers, but no research indicates what specific competencies can be developed at various grade levels or whether observation competence follows a developmental process.

The study has implications not only for the process of observation but for the science processes generally. If the viability of the Norris model and its usefulness as a criterion can be demonstrated using the proposed technique, then similar models and evaluations for other processes can be anticipated.

By developing a means of depicting the observation competence of the typical grade six student, it is possible to set goals for curriculum and instruction, and to evaluate the success of such curriculum and instruction in attaining these goals. The relative effectiveness of different curricula could be assessed against such a standard, and the effects of different curricula in promoting the process could be indicated more precisely.

Additional support for the necessity of promoting and evaluating observation ability comes from another direction, specifically, the field of critical thinking. In recent years, major efforts have been expended in defining critical thinking and in determining how it should be evaluated. Ennis (1962, 1980, 1985) has been a key figure in this field. His conception of critical thinking has been a major source of criteria for constructing tests of critical thinking. According to Ennis, one of the most important aspects of being a critical thinker is being a good observer. Observations are seen as being the basis for thinking critically.

Siegel (1980) has argued that students have a moral right to be taught how to think critically. For him, being a critical thinker, and consequently being a good observer, is of such significance and importance to students that not training them in this area is morally wrong. Knowing how competent our students are at observing gives some indication of their critical thinking ability. If there is a deficiency in this area, then, to pursue Siegel's line of reasoning, it is incumbent on us as educators to recognize that fact and to make some attempt at correcting the deficiency.

## General Overview of Approach

The basic approach used in this study is to interview individual students while they are conducting a series of simple science activities or experiments that require them to make and report observations. As they do this, the students are asked questions about the nature of their observations, and about their conceptual understandings that influenced their making such observations. Their answers reveal their thinking processes and the conceptual understandings that determine what observations are made and reported.

## Scope and Limitations

The study will not encompass the full extent of Norris's conception of observation competence. It will focus on student competence only in making and reporting observations. Determining students' ability to assess the reports of observations, which is also part of Norris's conception, would require an approach different from this study. Even with respect to making and reporting observations, however, there are certain competencies whose presence cannot be detected because of the nature of activities chosen for the study. For example, according to Norris, conditions that are conducive to observing well include the observer having no conflict of interest and the observer not allowing his or her emotions to interfere with making sound judgements. Norris also claims that in order to report observations well an observer should make the report himself or herself, close to the time of observing, and in the same environment in which the observation was made. The context in which the activities are set cannot be expected to yield information pertinent to such competencies. The Norris model defines observation competence in the broadest sense, but this study is conducted within the context of an elementary science classroom, so such factors as conflict of interest and emotional state are not expected to have an effect.

## Expected Outcomes

Among the important outcomes of this study is the demonstration of the power of an alternative means of assessing student competence with a science process. Whereas previous measures indicated to what extent students were able to use the process, this study indicates specifically what traits they possess or lack which influence their facility with the process of observation. Whereas previous measures have been typically paper-and-pencil tests, this study follows an interview approach whereby students actually demonstrate their observation competence within a scientific context.

As a part of the study, portrayals of individual students' observation ability are made. Specifically, portrayals of a below-average observer, an average observer, and an above-average observer are developed. Also, a "composite picture" of the typical grade six student is constructed.

## Summary

Acceptable methods for evaluating students' ability to use the science processes are not available. The problem is enhanced by the fact that the processes themselves have not been clearly defined and the development of tests using narrow definitions has resulted in the creation of inadequate measures. In particular, the process of observing has not been adequately defined, nor has there been an effective means of assessing the competencies that students possess with this process.

The need for an effective means of evaluating observation ability is supported not only in the science education field but also in the field of critical thinking. Because of the central role that competent observation plays in critical thinking, significant efforts have been expended in defining this process and in devising strategies for testing it.

This study demonstrates an alternate means to paper-and-pencil testing for evaluating observation ability. Criteria of being a good observer are proposed by Norris (1984) and will serve as the focus for this study. The model by Norris lists various conditions and principles that enhance observation competence and can guide evaluation for this process. Questions asked students are attempts to elicit information about their knowledge of these conditions and principles.

The end result is a detailed description of competencies students possess and lack that make them perform the way they do on observation tasks. This makes it possible to conceive of possible strategies of rectifying deficiencies. It also is possible to use the information acquired to set goals and to evaluate the relative effectiveness of various curricula and instruction in achieving those goals.

# CHAPTER 2: LITERATURE REVIEW

This literature review focuses on two questions considered fundamental to this study. In answering the first, "What is observation?", there is a description of the process of observation as it has come to be viewed in science and science education, followed by a discussion of observation as an important aspect of critical thinking ability. A discussion follows on the transferability of skills that is intended to integrate the two views of observation. In answering the second question, "How does one determine observation competence?", ability testing and some of the problems with reliability and validity are discussed. Some tests of observation ability are discussed and shown to be inadequate or inappropriate for the purpose of this study. A final discussion of interviewing procedures and techniques is presented.

## What Is Observation?

Good observation skills are crucial for the success of science. Observation skill has been promoted as a desirable outcome of studying science and, more recently, as an aspect of critical thinking. The concept of observation has been analyzed by philosophers and educators, and its role in science and science education has been a source of considerable debate for some time. This section discusses some of the literature that places observation at the fore of both science and critical thinking.

Gagné's Conception of the Science Processes

As early as the mid - 1800's science educators had argued that the processes of science should be taught as a part of school curriculum. In a brief historical account of the development of such ideas about science instruction, Finley quotes Layton (1973):

> In England, Thomas H. Huxley, Joseph D. Hooker, and John S. Henslow adopted the position that the unique characteristic of science as a branch of learning was the method by which knowledge was acquired and that the inductive aspects of scientific activity, rather than the conclusions, were of utmost significance from an educational point of view. Science was to be studied in the schools not for its informational benefits but because it trained the power of observation and reasoning. (Finley, 1983, p. 47)

Robert Gagné (1965) expressed a position similar to Huxley, Hooker, and Henslow in an address to the American Association for the Advancement of Science. He too believed that rather than teaching the content of science, much more emphasis should be placed on teaching students to become more proficient in the use of the science processes.

To understand what Gagné means by the science processes it is useful to see how he feels a person becomes an independent investigator, which is the ultimate aim for anyone intent on becoming a scientist. To become an independent investigator, Gagné believes that a person progresses through a series of levels or stages, with each successive one being dependent upon successful acquisition of those previous to it.

At the earliest level of instruction, the student needs to learn how to observe, how to measure, how to describe, how to classify, how to infer, and how to make conceptual models. According to Gagné, persons will use such capabilities all their lives. It is these and other processes which Gagné feels should form the basis of instruction in science through the first few years of school. By providing students an opportunity to practice these competencies in a wide variety of content areas we provide them with the skills needed to progress to the next level of being an independent investigator.

At the second level, the student uses previously learned skills to acquire a broad knowledge of principles of science through the various disciplines. This includes knowledge of content and method. Such learning forms the basis of science programs through the junior high school and into much of the senior high levels.

This broad knowledge of principles is required so that the student can move to the next level, the practice of inquiry. At this level, the student is able to speculate, to form and test hypotheses about scientific problems that are not trivial, and to self-criticize (Gagné, 1963, p. 151). Such inquiry should be practiced in the discussion class, in the laboratory, as well as in individual study. Gagné sees this level of instruction during the last year or so of high school and continuing into the first years of college.

This progression through the first three levels should have prepared the students for the fourth and final stage -- that of the independent scientist. They have mastered the process skills necessary to practice science, have broad knowledge in their own and other fields, and understand and have practiced inquiry such that they now know what they are doing. Students are able to begin a new line of investigation in a disciplined, responsible manner, with deliberate attention to what has gone before, but with minds unhampered by tradition (Gagné, 1963, p. 151).

Let's return to the first stage, that of practice with the processes of science. In keeping with Gagné's conception of a hierarchy of learning, the processes themselves have been organized from simple to complex with each skill fundamental to those above it. Stated from simple to complex, Gagné's list of processes are: observing, classifying, describing, communicating, measuring, recognizing and using spatial relationships, drawing conclusions, making operational definitions, formulating hypotheses, controlling variables, interpreting data, and experimenting. The processes are hypothesized by Gagné to be hierarchically organized such that the ability to use each upper level process is dependent on the ability to use the simpler underlying process. It is worthy to note that the process of observing is placed at the very bottom of the hierarchy, thus giving the impression that it is a very simple process and easily mastered.

This conception of the science processes has had considerable impact on research, curriculum development, and science education for more than twenty years. For example, science programs have been developed which focus on learning the processes as well as on emphasizing content. Programs such as Science - A Process Approach (AAAS, 1967-68), Elementary Science Curriculum Study (Crocker, 1973), and the Science Curriculum Improvement Study (1968), are such examples. A vast array of new research programs developed as science educators became interested in the process approach, and studies have been conducted to determine whether students also learn content along with the processes (Bredderman, 1982). In more contemporary programs, the balance between content and process is shifted back somewhat, but process learning still is a major aim in such programs as Addison-Wesley Science (1984), or Searching For Structure (1981).

<u>Observation -- A Fundamental Science Process</u>

According to Gagné's position, observation, by nature of its position at the foundation of the hierarchy of skills is considered to be the fundamental skill needed to acquire the broad knowledge needed to conduct inquiry. Gagné describes it as "the process of observing likenesses and differences in single objects that vary in their physical characteristics as detectable by any of the senses" (Gagné, 1965, p. 3). Such a view of observation has been presented to students through such programs as Science - A Process Approach (AAAS, 1967-68) and Elementary Science Curriculum Study (Crocker, 1973). In keeping with a major premise of Gagné's position, it is expected that the skill gained by observing in science activities would be transferable to other areas.

Gagné's view of observation has come under considerable attack by science educators and philosophers of education. Martin (1972) describes in detail the central role that observation plays in a scientific investigation and demonstrates how observation is more than detecting likenesses and differences or receiving sensory impressions. To illustrate this point, Martin asks his readers to consider the following statements: (a) Jones observed John's measles symptoms; and (b) Jones observed that John had measles. Martin points out that in the first statement Jones can observe the measles symptoms without knowing or believing that they have anything to do with measles. A visual image has been created, but the symptoms are not identified necessarily for what they are. It is quite possible that they may have been mistaken for something other than measles symptoms.

In the second statement, however, something more is implied. In addition to a visual image, the symptoms are identified as those of measles. Factors such as previously acquired information about measles symptoms are relevant in making such identification. Such claims as these made by Martin can be traced to the thinking of Hanson (1958). According to him, all observation is "theory-laden"; it is dependent upon interpretation using some theory, background information, or assumptions. The theoretical background, experience, and training of the observer greatly affects what is observed and can be observed. What can be observed is a function of this theoretical background and training.

This point is made clearer by Finley (1983). He challenges Gagné's position that, for example, the reception of light emanating from a geologist's polarizing microscope somehow results in knowledge. A geologist, for example, might observe sedimentary particles that are beginning to undergo metamorphism. A novice, or even a scientist from another discipline, on the other hand, might be unable to make such an observation although the same sensory stimulation has been received. At best, the novice observes a configuration of colours, shapes, and sizes. Again, background information and conceptual understanding play a fundamental role in determining what one can observe.

Of course one could counter that the novice and the expert are both making the same observations of colours, shapes, and sizes, but the geologist, because of background information and training, is able to make inferences from these that the novice cannot make. When one compares observation and inference in this way, the unspoken implication is that with the inference, but not with the observation, there is a possibility of doubt, and, therefore, of scepticism. But if one speaks of observation, this element of doubt is not present, or so it seems. Shapere (1982) has argued that this view is untenable. If propositions have been shown to be reliable in the past, there is no reason to continue to view them with doubt or scepticism merely because it is possible to have doubt or to be sceptical. From this line of argument it can be seen that, because of the reliability of the geologist's background beliefs, there is no reason to doubt the geologist who claims to observe sedimentary particles that are beginning to undergo metamorphism, even though the novice could observe only colours, shapes, and sizes.

Norris (1985) has provided a detailed philosophical analysis of observation in science and science education. He has argued that observations can be thought of as being on a continuum from simple to complex. Simple observations require very little background information and characterize the types of observations we normally have children make in the study of science. Such observations as noting colours, sizes, and amounts, could be considered simple and fit well with the view of Gagné. Complex observations, however, require very extensive background knowledge and often cannot be made with human senses. In fact, they may be outside the realm of human sensitivity altogether. Such observations as observing the bending of starlight as it passes the sun, or observing the centre of the sun by detecting neutrinos emanating from its core, fall on the complex end of the continuum.

Borrowing from Shapere's work, Norris also argues that the demarcation between observation and inference varies with the state of knowledge at the time. To quote Norris's example (p. 285): "at one time people might have thought that the far side of the moon was in principle not observable, but now it has been observed". What is observable changes with changes in human knowledge, or, as Shapere (1982, p. 492) says, "the specification of what counts as directly observed (observable), and therefore of what counts as an observation, is a function of the current state of physical knowledge, and can change with changes in that knowledge".

The preceding discussion of observation provides quite a contrast to the way in which Gagné and science education generally view observation. Gagné appears to believe that if a person has functioning senses, then that is all that is needed for knowledge to be acquired through use of those senses. Hanson, Martin, Norris, and Shapere, on the other hand, are saying that one's knowledge and beliefs determine the kinds of observations that one can make, and hence determine or guide the development of new knowledge. That is, previously acquired knowledge provides a context for making new observations. Observations are thus dependent upon the nature of the observer and the information the observer brings to the situation.

Observation -- An Aspect of Critical Thinking

An aim of contemporary science programs is to promote critical thinking. The importance of critical thinking was pointed out by Socrates (Glaser, 1985; Paul 1985) and more recently by Dewey (1933). Since the 1980's, critical thinking has been receiving an ever increasing amount of attention. Major efforts have been expended in defining just what the term critical thinking means, developing valid ways of teaching and testing for critical thinking ability, and in trying to determine whether critical thinking ability is a skill that is generalizable across disciplines or content domains.

In 1962 Robert Ennis published what has come to be a landmark paper in this area. He described his conception of critical thinking as one that "can serve as a basis for research in the teaching of and testing for critical thinking ability" (p. 83). In his conception of critical thinking, which he defined as the correct assessing of statements, Ennis lists twelve aspects which he believes, if followed, "may be looked upon as a list of specific ways to avoid the pitfalls of assessment" (p. 83). More recent work by Ennis (1980, 1985) has further developed and refined what he means by critical thinking. He has expanded his definition so that he now maintains: "Critical thinking is reflective and reasonable thinking that is focused on deciding what to believe or do" (Ennis, 1985, p. 45).

In critical thinking as conceived by Ennis, observation serves as a basis for making inferences and deciding what to believe or do. Observations may come as information from others or through making one's own observations (Ennis, 1985). The information thus obtained is then used to infer to a conclusion -- a decision about belief or action.

Judging whether an observation statement is reliable is one important aspect of deciding what to believe or do. To help clarify and define what leads to good observations, Ennis lists a set of conditions that would assist a person in making and reporting observations. He also specifies a set of principles for assessing reports of observations.

Knowing the extent to which the conditions have been met and then following the principles of assessment will assist a person in deciding whether an observation statement is reliable.

Norris (1984) has further developed and refined the principles and conditions that Ennis presented. Norris presents a set of conditions and principles that facilitates making and reporting observations as well as assessing the reports of observations. Norris contends that the critical thinking proficiency of taking into account these conditions and principles is what is needed to observe well in science. Thus, Norris provides a link between observation as an aspect of critical thinking and observation as a science process. Therefore, much of the research on critical thinking has direct relevance to observation as a science process.

In addition to proficiencies, recent writings have focused on critical thinking dispositions. For example, Sternberg (1983) claims that no matter what level of critical thinking skill a person possesses, it is of no practical benefit unless the person is disposed to use these skills when they are appropriate. As well, Ennis (1985) includes the following dispositions in his conception of critical thinking: being open-minded, paying attention to the total situation, seeking reasons, and trying to be well-informed. The same can be said for observation ability.

The fact that someone has the ability to observe well is no guarantee that they will use it when appropriate. Having the ability to observe well is of little value if one is not disposed to use that skill.

The ability to think critically is being promoted as a goal of education and science education in particular. This is likely to become an even more important goal as knowledge continues to accumulate at an increasing rate. When observation is seen as a basis for critical thinking, then it is evident that observation ability needs to be promoted and effectively evaluated.

Transferability of Skills

An important claim made by Gagné (1963) is that the science processes are generalizable across content domains and contribute to rational thinking in everyday affairs. This claim has met with mixed support as can be seen from what follows. For example, Finley (1983, p. 53) argues that the "processes will be different from discipline to discipline and different even within a discipline when different conceptual aspects of the discipline are in use. ...it is unlikely that there will be content-free intellectual skills that are generalizable across multiple enquiries." Norris, on the other hand, would seem to support Gagné's belief that skill at observing is a transferable skill applicable to all disciplines, as evidenced by the following statement: "Although observational competence is needed in many fields and in many aspects of everyday living, science is a very sensible place to promote it" (Norris, 1984, p. 141). The same divergence of view exists in the critical thinking field. Some believe critical thinking is transferable (Ennis, 1980, 1985, 1989; Norris, 1984; Norris & Ennis, 1989; Paul, 1985; Perkins & Salomon, 1988, 1989; Sternberg, 1987) while acknowledging the importance of local knowledge on which to exercise those skills. However, the view that there are general critical thinking skills that are transferable across disciplines or domains has met with considerable opposition. One of the strongest critics has been McPeck (1981). Furthermore, Schoenfeld (1985) and Hirsch (1987) argue that content-specific knowledge, not general problem solving heuristics, is the key to success in dealing with situations. Their belief seems to be that if we provide students with a broad knowledge base about their culture, then their thinking skills, the ability to apply their knowledge, will develop naturally.

Much remains to be resolved in this area. However, if skill at observing is indeed general, there is much to be said for promoting it in our schools and for finding effective ways to evaluate it. When conceived as a general skill, observing assumes a vital role not only in the science classrooms, but in all aspects of daily life. Having good observation skill is a precondition to being a good critical thinker, and being effective thinkers in the information age is believed by many to be the alternative to being swamped by an overabundance of subject matter. Viewed in this way, there is little doubt that observation skill needs to be effectively taught and evaluated.

Synthesis and Evaluation

From this section of the literature review, certain ideas can be extracted that can guide the evaluation of observation ability. Of special importance, is the view that observation is not as simplistic as Gagné presents it. Knowledge doesn't come about just as a result of sensory stimulation. Instead, making accurate observations requires a background of related knowledge that can be brought to bear. If one were to assess observation ability on the basis of what Gagné believes about the process, then one would need to determine if the subject can report things accurately in accordance with what an experienced investigator would agree as being correct. But according to the more detailed philosophical ideas of such people as Ennis, Hanson, Martin, and Norris, assessing observation ability requires one to get at what the subject already knows and how that is applied in a given context of making observations. Merely knowing whether or not a subject reported an observation correctly is not enough. One needs to gather information related to the sorts of things the subject heeded, the precision that was used, the extent to which the subject was alert, and other such factors. These are the kinds of things that Norris (1984) believes determine observation skill, so, if we wish to assess ability with this skill, we need to get close to what is actually taking place within the subject's mind. We need to assess the process itself and not just the products of that process.

Given this understanding of observation and how it needs to be assessed, we need to conceive of a method by which to learn about the knowledge a subject is heeding as observations are made. Science and technology have not progressed to the point where we can look into a person's head and analyze the electrical impulses that are flowing through the circuitry of the brain, although that might be a possibility for some remote future time. We are left with having to conceive of some other method by which to gain access to what is going on in a person's mind, so that inferences can be made about their observation ability in general. In the next section, I offer suggestions on how it may be possible to get at that kind of knowledge.

## Determining Observation Competence

This section discusses attempts to measure observation ability both as a science process and as an aspect of critical thinking. I shall look first at the field of ability testing in order to point out certain criteria for judging the observation tests. In particular, construct validity and the types of evidence that are needed to support claims for validity are examined. Special reference is made to the validity of interviewing procedures that can be incorporated into a methodology for assessing observation ability.

Ability Testing

Reliability and validity are the two main criteria that have to be met when developing ability tests. In what follows, each of these is discussed and their relationship to the present study is shown.

Reliability. Reliability refers to the extent to which a test yields consistent results from one occasion to another. A person may perform differently on one occasion than on another for reasons that may or may not be related to the purpose of measurement. A person may try harder, be more tired or anxious, have greater familiarity with the content of questions on one test form than another, or simply guess correctly on more questions on one occasion than on another. For reasons such as these a person's score will not be perfectly consistent from one occasion to the next.

A perfectly reliable test for measuring observation skill would yield the same results when given to the same individual at closely spaced time intervals. The extent to which two such administrations of a test correlate with each other is expressed as a reliability coefficient.

A difficulty with this method of rating reliability has been noted. Not only do we get errors of measurement caused by factors such as fatigue, anxiety, and emotional state, there is the added factor that characteristics of people can be changed by the very act of trying to study those characteristics. To get around this problem, evaluators have devised means of estimating the reliability of an instrument from a single administration. One such method, called the split-half method, requires splitting the items on the test into halves, and then correlating the scores on the two halves. Another method of estimating reliability with a single administration of an information-gathering instrument is the Kuder-Richardson method. This procedure involves intercorrelating all the items in all possible combinations and computing an average correlation. To the extent that such correlations are high, a test can be said to be reliable.

A test of observation, however, that tested for various aspects of observation ability might be disadvantaged by the Kuder-Richardson method of computing reliability. As Norris and Ennis (1989) pointed out, to the extent that observation ability is heterogeneous, a high correlation necessarily need not be expected among items testing different aspects. In fact, a low correlation could be presented as evidence for the heterogeneity of observation ability and of the quality of the test. Thus, reliability coefficients must be interpreted carefully.

Because of the format of some tests, they cannot be sensibly divided into units. For example, some tests of critical thinking follow an essay format, or some tests are conducted orally whereby a subject responds to probing questions. In either case, to divide such tests into units is difficult, if not illogical, and makes the Kuder-Richardson method of determining reliability unsuitable. A technique frequently used with such tests is called inter-rater reliability. In this procedure the ratings of different judges, or of the same judge at different times, for a single administration of a test are correlated. The reliability thus obtained gives an indication of the consistency of scoring. However, the original concern of reliability is with consistency of student performance, and this method tells us nothing about that. Judges' performance in scoring is not at all the same as student performance on repeated administrations of a test, and therefore cannot replace estimates of reliability in the original sense (Norris and Ennis, 1989, p. 48).

Reliability estimates provide assurance that a test will yield similar scores when given to the same individual on more than one occasion. However, in addition to reliability, we need some assurance that a test is measuring what we believe it to be measuring. Standards of validity are required.

Validity. According to the APA Standards (1985) "validity refers to the degree to which evidence supports the inferences that are made from the scores" (p. 9). The types of evidence for validity are categorized as content-related, criterion-related, and construct-related. Content-related evidence refers to the degree that the sample of items, tasks, or questions on a test are representative of some defined universe or domain of content. The evidence consists of the judgement of experts in a subject who can confirm that the test is representative of the field in question. In this study, content validity is determined by the extent to which the definition of observation conditions and principles is complete and the extent to which the tasks given students sample them.

Criterion-related evidence refers to the extent to which the test scores are systematically related to one or more outcome criteria. Basically, this means that a new test is correlated with previously acceptable measures and the pattern of correlations studied. A test should correlate highly with those measures it is logically related to and correlate lowly with other measures. The value of a criterion-related study thus depends on the relevance of the criterion measures that are used (APA, 1985, p. 11). In the present study criterion-related evidence consisted in the relationship of the children's performance to their school science grades.

The third type of evidence, construct-related, was given serious attention and consideration by Cronbach and Meehl (1955). According to them, construct validation is involved whenever a test is to be interpreted as a measure of some attribute or quality that is not "operationally defined". In such an interpretation, an investigator is trying to describe "What constructs account for variance in test performance" (Cronbach and Meehl, 1955, p. 284). A construct is a postulated attribute of people, assumed to be reflected in test performance. In test validation, the attribute about which we make statements in interpreting a test is a construct.

According to Messick (1975) all measurement should be construct referenced. A measure estimates how much of something an individual displays or possesses. The basic question is "What is the nature of that something? It may be answered by referring to evidence in support of particular attributes, processes, or traits construed to underlie and determine task performance" (p. 957).

It would appear from such comments that constructs can be rather vague concepts, difficult to define or conceptualize, and often not obviously connected or related to other better understood constructs. Nevertheless, if one is seeking to design a test to measure the extent to which a person does possess a certain trait, ability, or construct, then one does need a good understanding of what that construct is and what would count as evidence that it is present to a certain extent.

Cronbach (1971) refers to the need of constructs being defined. According to him, "Construct validation begins with the claim that a given test measures a certain construct" and that "[t]his claim is meaningless until the construct is amplified from a label into a set of sentences" (p. 47). He seems to be suggesting that, by collecting the meaningful descriptive sentences that become associated with a construct, we are in essence building a set of criteria for the construct. This can then be used to determine if the test in fact matches the criteria, or ideally, it can be used as an aid in test construction. This brings us into the shadowy area of content-construct confusion of which Mary Tenopyr (1977) writes, for by defining a construct precisely, we are also more nearly defining a universe from which to draw items for a test. In this context then, content-related evidence could be taken as one form of construct-related evidence.

The evidence classed in the construct-related category focuses primarily on the test score as a measure of the psychological characteristic of interest. The construct of interest for a particular test should be embedded in a conceptual framework, no matter how imperfect that framework may be (APA, 1985, p. 9). Typically, the process of compiling construct-related evidence for test validity starts with test development and continues until the pattern of empirical relationships between test scores and other variables clearly indicates the meaning of the test score.

The measure of intelligence, as reported by Stanley and Hopkins (1981), provides a classic example of validation. Early attempts to measure intelligence using reaction time, auditory memory, and other psychomotor and psychological measures were discarded, because performance on these measures did not correlate with other behavioural evidence of intelligence, such as school grades. The expected and logical relationships between relevant variables were not confirmed, that is, the tests lacked criterion-related evidence. Later, the French physician Alfred Binet constructed tasks that were logically related to intelligence; they required complex cognitive abilities. Many of Binet's tasks were found to be related to other variables in a manner expected of a measure of intelligence.

Gradually, through a continual process of research and revision, these tests yielded scores that agreed substantially with logical and theoretical expectations. For example: (a) the scores correlated with age until maturity and then levelled off; (b) the scores had a substantial correlation with academic achievement;

(c) children who repeated grades scored much lower on these tests than those who were promoted; (d) the I.Q. scores yielded by these tests showed some stability over a period of years; (e) persons with clinical types of subnormality, such as mongolism, performed poorly on the tests; and (f) the correlation of identical twins was extremely high -- much higher than for fraternal twins. Such information as this illustrates the incremental procedure inherent in validation (Stanley & Hopkins, 1981, p. 105). In a sense, every bit of information about a test has relevance for validity, that is, in establishing what it does and does not measure.

Observation competence can certainly be construed as a construct. It is clearly a mental ability and cannot be directly observed, hence the only way we can learn about it, at least for the present, is indirectly, through inferences we make based on what is observable. It has been closely analyzed and given meaning by Ennis (1962, 1980) and Norris (1984). A set of conditions and principles for being a good observer has been the result of the analysis which they have put forth. The task then for validation of a testing method is to show whether the procedure being used does in fact test for those specific competencies.

In this study the intent is to gain information about what is happening in the subjects' heads as they observe. Thus, it is necessary to find a method that allows inferences to be made about the kinds of thinking that take place as observations are made. Some tests use an essay format whereby students display their ability to think critically by responding in essay form to various situations. Rather than use an essay format, however, the present study has students respond orally to situations. This allows for interaction between the subject and the investigator and allows for the clarification of ambiguous statements. In effect it is an attempt to bring the investigator a "step closer" to what is happening in the subjects' heads and goes beyond the barriers of a written test. Ericsson and Simon (1980, 1984) have presented very useful guidelines for investigators using such a research method. However, in order to justify an interview procedure as a valid method of testing for observation competence, it is necessary to discuss this methodology in some detail to show how it has been applied in other studies and how it might be adapted to the present study.

Interviewing Procedures

One method used to gain information about the course and mechanisms of cognitive processes is to ask subjects to relate verbally their thoughts. As early as 1917, E.L. Thorndike focused attention on subjects' verbal reasoning when reading as a means of understanding the nature of comprehension. On account of this and more recent endorsement (Thorndike, 1971), there seems more recently to be a growing trend towards this type of approach to the study of thinking.

Many studies of thinking processes have been conducted by use of protocol analysis. These will not be reviewed here. Instead, I shall discuss a main area of contention -- that is whether verbal reports can provide useful data.

Nisbett and Wilson (1977) argued that people do not have access to cognitive processes that cause behaviour. Instead, they claim that people often cannot report accurately on the causes of their thinking. When asked for such causes people do not interrogate their memories, but provide a hypothesis of what might have been the cause. As a result subjective reports about higher mental processes are sometimes correct, but when they are it is not due to direct introspective awareness. Instead, it is due to the incidental use of the right hypothesis. The implication seems to be that if you want to know why a person thought something, you may as well ask someone else as that person!

Such claims are disconcerting to anyone interested in conducting research through analysis of verbal data, because, if true, they discredit completely the value of any data received through such means. However, Nisbett and Wilson do suggest situations in which accurate verbal reports can be expected. These are characterized by "an available influential stimulus, which is a plausible cause of the response, and a lack of other plausible causes of the response" (p. 253).

Smith and Miller (1978) criticize Nisbett's and Wilson's conclusions for being too sweeping. They argue that it is possible to gain access to mental processes. Although they admit that there are situations in which access to processes is not possible, they maintain there are other situations where reliable access is possible. Thus, Smith and Miller claim that research should focus not on the question of whether people have access to mental processes, but on the question of the conditions of such access.

The position that people do have access to their thinking processes was further argued by Ericsson and Simon (1980, 1984) in developing a model of how subjects verbalize information from their short-term memories in response to instructions to think aloud. The central point of the Ericsson and Simon theory is that "information recently acquired (attended to) by individuals is kept in short-term memory and is directly accessible for further processing (eg. for producing verbal reports)" (p. 223). The act of verbalization is predicted to have no effects on the course of cognitive processing, but may slow down the speed of task performance. When the information is not available in short-term memory, it must be retrieved from long term memory. The information that can be recalled depends on the nature of the cues and probes provided.

A very significant aspect of the Ericsson and Simon theory is that it predicts the trustworthiness of verbal reports. The less leading the probe employed, the more accurate the information obtained, and the more leading the probe, the less accurate the information obtained will be. This idea of leading and non-leading probes can be clarified through example. If, in the course of probing, the interviewer says, "Would you say more about that?" or "What do you mean by that?" these could be considered non-leading probes because they do not provide the subject with any information except a request to elaborate. But, if the interviewer were to ask, "Does X make any difference?", where X is some specific piece of information, this would be considered a leading probe because it contains information that may not have been noted by the subject without the probe.

Norris (1990) reports a study of interviewing effects that was carried out in the course of validating a test on appraising observations. In the study some students who were taking the test thought aloud as they worked through items; some were asked non-leading questions, others were asked leading questions about decisions they had made on the test. Comparisons were made of how students performed across the various groups. Finding no performance differences, Norris concluded that "the elicitation of verbal reports of thinking did not alter subjects' performance and, by inference, did not alter their thinking [from a non-interview format]" (p. 47). These findings are encouraging for those who would use verbal data for conducting research.

Perkins (1981) has written about the use of verbal reporting as a means of gathering information about the thinking process. According to him there is an art to helping people to "share their minds" with an investigator. Of special significance are his remarks regarding retrospection. He suggests that a good way to obtain retrospective reports is to ask people to think aloud. Also, he suggests that one might set problems that are solved silently and ask for a retrospection immediately after an answer is given. "Can you tell me (or write down) what thought led to the next over the last several seconds up to the point you have just arrived at? Try to indicate what happened step by step, but only report what you actually remember now, not what you think might plausibly have happened" (Perkins, 1981, p. 37).

The instruction and follow-up questions for a retrospective report encourage remembering rather than reconstructing, according to Perkins. He predicts that with retrospectives, more so than with thinking aloud, people tend to offer plausible explanations. Such explanations, according to Perkins, weave together their bits of memory to provide a fully coherent and motivated account of what happened. "Making sense of the record is the business of the investigator, not the subject" (p. 37).

Finley (1986) reports on the use of clinical interviews to complement the information that is received in a testing situation. According to Finley, researchers frequently attempt to find out the changes in knowledge that result from instruction. Clinical interview information can be used to determine the importance of the knowledge differences indicated by the test scores, and to make specific suggestions for instructional improvement. The information can also help determine what the test measured.

Two conclusions that Finley reached regarding the use of clinical interviews are of interest: "(a) Clinical interview results provide information that is not available through the use of the more typical tests, and (b) Clinical interview results provide insight into how the more typical achievement tests probably had functioned" (Finley, 1986, p. 648). Such conclusions are valuable because they tend to indicate that through clinical interviews it is possible to gain more information about students' conceptual understandings than can be gained from paper and pencil tests. Furthermore, the interviews lead to a better understanding of why students performed as they did on the related tests.

Larkin and Rainard (1984) suggest that any questions or probes that are intended to elicit explanatory responses be as non-leading as possible. The questions should merely echo examinees' reported thoughts or ask the subject to explain a little more what was said. This method of using non-leading probes is advocated by others (Afflerbach & Johnston, 1984; Ericsson & Simon, 1980, 1984; Loftus, 1979; Norris, 1988) who support the use of protocols in research.

The research on eyewitness testimony is another area that has relevance to the proposed study. Eyewitness testimony is often contained in verbal reports of what people can remember, or claim to remember. Research on eyewitness testimony describes the effect of different types of questioning on the accuracy of reports. Three categories of questions have been studied: (a) those eliciting free reports, (b) those eliciting controlled reports, and (c) those eliciting alternate-choice reports (Loftus, 1979, p. 90).

Research by investigators such as Clifford and Scott (1978), Dale, Loftus and Rathbun (1978), Lipton (1977) and Marquis, Marshall and Oskamp (1972) on the influence of such types of questions has provided three useful conclusions. Free reports tend to be more accurate than any other type of report but contain the least amount of information. Controlled reports are somewhat less accurate but provide relatively more information. Alternate-choice reports have the lowest degree of accuracy but contain the greatest amount of information. These conclusions are essentially in accord with the predictions that Ericsson and Simon (1980, 1984) make about the trustworthiness of verbal reports.

It can now be suggested that the literature and research on verbal reports as data, eyewitness testimony, and interviewing techniques have produced promising results for using an interview format to study the observation ability of students. Ericsson and Simon have theorized that information recently acquired or attended to by an individual is kept in short-term memory where it is available for further processing, such as producing verbal reports. This would imply that when students are making observations, the information they are receiving, and the background knowledge that they are drawing on, are held in short-term memory and are therefore available for reporting. The task then for an investigator who wants to get at this information is to get the subject to verbalize about the contents of short-term memory without altering the course of the subject's thinking from what it would have been had the subject not verbalized.

The preceding literature suggests that, in eliciting reports of thinking, a number of techniques can be used. These include having the subject think aloud, whereby the subject is expected to verbalize the thoughts that are going through the mind. One could also probe the subject with non-leading probes about what has been verbalized and thus encourage the subject to say more. So, for example, if the subject reports an observation, or says something about what was done or observed, the investigator could pose a question or give a direction to have the subject elaborate. Leading questions can also be asked although the eyewitness testimony research indicates that the information contained in the responses tends to be less accurate than that elicited by non-leading probes. Retrospective reports can also be requested whereby subjects are asked to report on how and what they had thought as they performed an activity or made observations.

Norris's finding, that the elicitation of verbal reports does not alter the course of subjects' thinking from what it would have been in a non-interview format, was arrived at in a study utilizing a test of observation appraisal. If we generalize the finding to testing the ability to make and report observations, then the thinking that takes place using an interview format can be expected to be the same as would have occurred had there been no interviewing. Furthermore, the eliciting of the reports does not alter the course of students' thinking as subsequent observation tasks are performed. The information thus gathered can, according to Finley, be used to create an understanding of why students observed as they did.

Testing Observational Ability

The elementary science curricula of the 60's placed less emphasis upon scientific information for its own sake and more upon the processes by which this information was discovered. New techniques of evaluation, particularly tests of process development, were eagerly sought. Precise definitions and measurement of science skills were of utmost concern to science educators.

The following twenty-five years produced numerous instruments whose purpose was to measure student competencies with the processes generally or with selected processes. The literature abounds with instruments specially designed for graduate studies and academic research. However, for the purposes of the present study, the instruments have two important deficiencies. The first relates to how poorly the science processes, particularly observation, had been defined. This meant there was little guidance for judging what the tests were designed to measure. The previous discussion on validity dealt with the importance of having well-defined constructs when designing tests of mental abilities. The following discussion will illustrate that, except in a couple of instances, there was no clearly defined construct that led to the design of the tests that purport to measure observation ability.

A second deficiency with the tests of observation ability is summed up in this comment from Ebel (1972): "The price that must be paid for a test's advantages of efficiency and control in the observation of student achievement is some loss in the naturalness of the behavior involved" (p. 13). To clarify what Ebel means by "naturalness", consider the following situation. An investigator may want to evaluate the skill that a subject has in bicycle riding. The most logical approach is to have the subject ride a bike while the investigator makes observations of the behaviours that the subject exhibits. Skills such as balance and co-ordination will be employed by the subject and the investigator can directly observe the extent to which the subject displays those skills in a rather natural setting. Special note can be made of the subject's strengths and weaknesses.

On the other hand, if the investigator needed to assess the skill of many subjects, it may not be feasible to analyze individual performances due to the extensive time required in following and observing each bicycle rider. The investigator is faced with the task of developing an alternate, efficient means of assessing the level of skill in the group as well as in individuals.

Let's assume, for illustrative purposes, that there is a correlation between bicycle riding skill and knowledge of bicycle parts and safety rules. The investigator may choose to design a paper-and-pencil test which questions subjects about their knowledge in these areas. From the obtained scores the investigator infers the ability of the group as a whole and the relative ability of individuals. However, the scores are not useful in detailing which strengths and weaknesses exist within the group for the test reveals nothing about skills such as balance or coordination. The investigator has an efficient means of quickly assessing a subject or a group but much has been lost in the naturalness of the behaviour.

Many of the behaviours that we may wish to evaluate cannot be so easily or directly observed as bicycle riding skill. Skill at observing is one such example. Observing is an activity that takes place in the mind of the observer and this makes it an unobservable. If we wish to get some measure of the observing ability that a subject possesses we will have to resort to less natural settings and then infer to the ability. For example, one way to do this is to have the subject take a multiple choice paper-and-pencil test whereby the subject is asked to make observations in items on the test and select an answer. The extent to which the student's responses match the keyed responses gives the investigator an indication of how well the subject performed. The investigator has an efficient method and tight control but the observing act is not taking place in a natural scientific setting. The investigator can make but limited inferences about the precise nature of the subject's observation ability.

Another method that the investigator might use is to have the subject actually make observations of real objects. For example, the student may perform an experiment and report the observations to the investigator. This setting is somewhat more natural than the paper-and-pencil format, however, some efficiency and control has been lost, because now the subject determines the course of the activity and the choice of response.

A perfectly natural setting would be one in which the subject goes about daily activity and, unknown to him or her, an investigator somehow has direct access to what is going on in the person's brain. The investigator would be able to interpret the interaction of electrical impulses to understand exactly how the subject processes information and just what information is processed as observations are made. Such a procedure is of course impossible.

Somewhere along this continuum from unnatural to natural settings is where we will have to settle for gathering information about people's observation ability. If we administer a paper-and-pencil test, we can learn how successfully a subject has observed compared to the rest of a group that has taken the same test. But if we wish to learn specifically which particular observation skills the subject possesses or lacks, then, as in the case of evaluating bicycle riding skill, we need to get as close as possible to the actual activity so that we have the best possible access to the process itself.

In light of the two deficiencies just mentioned, poorly defined constructs and unnatural settings, a number of tests will be discussed to show that there are no suitable ones available for gathering the types of information being sought in this study. It will therefore be necessary to design a method that can be used to get closer to what does happen in subjects' minds as they make observations.

<u>Observation Tests</u>

Hungerford and Miles (1969) designed a paper-and-pencil test to measure the observation and comparison skills of junior high school students. The particular subject matter selected for use in the test was deciduous winter twigs. Students were required to make an accurate visual reproduction of a particular specimen. Visible morphological details were to be included. Students were to label their drawings with technical vocabulary if it was familiar to them or their own vocabulary if they were not familiar with the technical vocabulary. The excellence of the drawings and the labelling of the important morphological features were criteria used in scoring the test. For this test, observation is defined as "the ability to make accurate observations with the subsequent communication of these same observations. The communication of observations can be oral, written, visual, or a combination of these modes" (p. 62).

To illustrate the weakness of this test, let's assume that, having taken it, a student is found to have poor observation ability. A teacher is faced with helping the student to improve. Where is the place to begin? The test score tells merely the extent to which the student observed successfully. It is not helpful in specifying the competencies the student lacks. For example, maybe the student observed poorly because he or she did not exercise care, was not alert, or did not use a precise technique. On the other hand, maybe the student made the observation well but reported it poorly because he or she did not know how to keep adequate records. Such factors cannot be determined from this test for it was not designed to give such information. Furthermore, the term "observation" is too narrowly defined by Hungerford and Miles to allow construction of a test that would yield such information, and consequently this test cannot be considered adequate for measuring other than the rough magnitude of ability.

Somers and Lagdamen (1975) devised an instrument to measure the ability of children to observe, compare, and classify geometric figures. Students were to use these processes to "detect similarities and differences" in characteristics of geometric figures or sets of figures. Circles, squares, rectangles, trapezoids, and isosceles triangles were utilized. The figures appeared in either dark blue, yellow, light blue, or red, and each was made in three sizes, all having proportional dimensions. Sample items include:

A. Which object has the most number of sides?

a. triangle 8

b. trapezoid 11

c. triangle 11

B. Which object is most similar to this object (triangle 10)?

a. square 6

b. triangle 8

c. triangle 6

One could, no doubt, question the extent to which such items are really observation items. But there is another concern that exemplifies a general problem with such tests. This test was of the paper-and-pencil variety and again we have a recurring problem. While it was assumed that choosing the keyed answers indicated that the students used the processes correctly, such choices do not reveal what led the person to choose the keyed answers. It does not reflect the thinking process that goes into making decisions on the test. The test does not reveal particular competencies the student may or may not possess, because it is not based on precise analysis of proficiencies that facilitate a person's ability to use the processes concerned. Hence, while the test provides tasks that supposedly require the subject to use the process of observing, the single score will tell only to what extent the student has observed successfully. It will not tell whether the individual possessed certain competencies such as alertness or good conceptual understanding. Again, the observation construct is narrowly defined and does not guide the test designer in selecting items that may indicate the presence of specific competencies.

Science -- A Process Approach, developed by the American Association for the Advancement of Science and Xerox Corporation (AAAS, 1967-68), provided competency measures at the completion of each unit of work with a particular process. The purpose of such measures was to determine whether children learned what the exercise was intended to teach -- that is, a particular process skill. For example, at the end of a unit that promoted the skill of observing, the following competency measure is found in SAPA, Part B, Observing 11, 9th page:

> Give the child six iron objects of different sizes and shapes (two magnets and four non-magnets). Give the child a box of paper clips. Say, show me how you would find out which of these objects are magnets and which are not magnets. Put one check in the acceptable column for Task 2 if the child attempts to pick up the paper clip with each of the metal objects. Put one check in the acceptable column for Task 3 if the child identifies the two magnets.

Here we see a substantially different approach from that taken in the typical paper-and-pencil test. The student is expected to speak and to manipulate objects to demonstrate learning. This is somewhat removed from the abstractness of paper-and-pencil tests and can be viewed as more natural. However, a problem occurs, for again the process has not been clearly defined and, except for intuition, the teacher has little means of deciding what would count as evidence that the student possesses or lacks competence in making observations or what special proficiencies facilitate the making of accurate observations. As with the paper-and-pencil type tests, it is still the products of the student's observing activity, and not the process itself, which are of concern here. It will reveal little about a student's observation ability other than whether he or she observed successfully for a given activity. This is of little value if we want to take corrective action to improve a student's observation ability or to make adjustments to a program to better promote the acquisition of observation ability generally, because it does not reveal specifically in which areas the students have strengths or weaknesses that contribute to observed level of performance.

Nelson and Abraham (1973) developed a procedure for measuring skill in observing, making inferences, verification, and classification. This was not a paper-and-pencil test as most tests of process skills have typically been. For this test, observation was defined as the ability to gather data through the use of the five senses. The test is described as follows:

> A sealed box with a number of colored sticks protruding from it is placed on a table in front of a child. He is told to examine the outside of the box using all of his five senses, and to tell as much about the outside of the box as he can. The person administering the test records verbatim the child's statements. No attempt is made by the tester to prompt or clarify the testee's remarks. (p. 291)

In interpreting the data, the rules for judging whether something is an observation statement are: (a) Statements are to be made about the outside of the box; (b) Statements describing sounds coming from the inside of the box will be considered further; and (c) All other statements about the inside of the box are considered incorrect.

A problem inherent in this test is the simplistic way in which observation is viewed. It seems to be following the lead of Gagné in which observations are viewed solely as information received through the senses. It does not consider the role of the background knowledge of the observer as current writers claim it should. The test designers also draw a sharp line between observation and inference. Except for observing sounds from inside the box, no other statements about the inside of the box can be considered as observations. For example, if the sounds of a local radio station are emanating from the box and giving information such as local news and correct time and date, the observer cannot observe that there is a radio inside the box, but can only report that there is a voice coming from inside the box. Reports of what the voice actually says are also acceptable. This conception of observation is not compatible with the way observation is conceived in current literature (e.g., Norris, 1985).

The other problems that confront tests such as this are also present. The process of observation is very narrowly defined and does not specify what competencies lead to good observing. Also, although the test is more natural than the abstractness of paper-and-pencil tests, it is still the results of observing, not the process itself, which are being evaluated.

I have been unable to find in the literature a test solely for measuring observation ability as a science process. Like the ones previously described, it is possible to find tests that attempt to measure observation skill along with other processes. However, they have been developed with a narrow definition of observation and while the pertinent test items may require observation ability, there is no specification of what particular competencies are applicable to items on the tests.

In the field of critical thinking there is a similar shortage of instruments for measuring observation ability. However, there does seem to be greater attention to the necessity of providing well defined constructs. The Cornell Critical Thinking Test, Level X (Ennis & Millman, 1985) has a section that tests for observation ability. The aspect of observation ability which the test measures is the ability to assess observation reports, not the ability to make observations, although the former may assist with the latter. The test is intended primarily for junior and senior high school and first year college, but has been used in grades four through six. What is significant, however, is the extent to which the construct is defined and the extent to which the items on the test are selected to reflect knowledge of the principles of good critical thinking. But the test is broad, measuring critical thinking ability in general; observation ability is but one aspect. In order to get detailed knowledge of observation ability it would be necessary to design an aspect specific test.

Norris and King (1984) designed a critical thinking test on appraising observations. The test was developed so that questions were based on a set of principles that "catalogue the effects which such factors as conflict of interest, degree of observational access, adequacy of technique employed, and extent of independent corroboration, have on the trustworthiness of what people claim to have observed" (Norris & King, 1984, p. 7). Though the test was designed to assess competence in only one aspect of being a good observer, it does have implications for assessment of other aspects, those of making observations well and reporting them well (Norris, 1984). Equally significant was the method used to determine construct-related evidence for the test validity. Students were interviewed as they did the test. The intent of this interviewing was to get some insight into the kind of thinking that went into making decisions on the test. It was possible to determine whether students were considering the principles, which Norris (1984) describes, for thinking about the various test items. Changes could then be made in the development stages so that good performance could be explained by examinees' following sound thinking processes, and poor performance could be explained by deviations from such processes (Norris, 1990, pp. 55-6).

While the Norris and King test is not suitable for assessing the ability of grade six students to make observations, it does offer suggestions. Designing an instrument around such a criterion and validating it by the method used offers suggestions that may be incorporated into process measurement. Specifically, the method can, with some modification, be extended for assessing competence with making and reporting observations at the grade six level. Such suggestions will be spelled out in detail in the next chapter.

In summary, it can be stated that there seems to be no acceptable method or test for assessing the observation ability of grade six students. While there are tests available whose designers claim they measure the extent to which a student has made successful or correct observations, none of the available tests will indicate what has led students to perform as they have, and it is this information that I need for this study.

## Summary and Conclusions

The discussion in this chapter has shown that observation is a fundamental science process and is also fundamental for critical thinking. The literature on transferability of skills has led to the conclusion that whether students are observing in school science activities or trying to make a decision about what to believe or do, the observation skills that are employed in both situations are essentially the same. Observation skill, it was argued, is considered so important in all aspects of a person's life that it is considered worthy of being taught in school. Not to do so is a disservice to students and is morally wrong.

Since observation ability is so important, it follows that we need effective means of evaluating for its presence.

We need to know how effective we are in teaching this skill so that we can make adjustments to compensate if we find that an inadequate job is being done. The literature on testing has shown that any testing device must meet standards of reliability and validity. The test must yield consistent results from one occasion to another for the same student, and the construct must be defined clearly so that we know what we are trying to measure and that the tasks on the test are clearly related to the construct. This means that if we are to devise a method of testing observing ability, a first requirement is that we know precisely what observation competence is. We need a set of criteria that outlines conditions that facilitate good observing. The extent to which a subject meets those conditions gives us an understanding of how skilled the subject is with the construct and where, specifically, the subject has strengths and weaknesses.

Previous tests of observation ability have been plagued by the fact that the construct was poorly defined. Therefore, such tests are unsuitable for specifying what particular skills or abilities a subject possesses that result in the perceived level of observing ability. The literature indicates that most observation tests are of the paper-and-pencil variety although there are some tests that require students to make observations of real objects or events. In all of the tests, however, the focus is the products of observing and not the process itself.

Assessing the process of observing requires that we have as natural a setting as possible so that the subject can make observations unimpeded by the artificiality of a paper-and-pencil test. Although artificiality cannot be eliminated entirely, the extent to which it is eliminated determines the naturalness of the testing situation. Our desire then is to understand how the subject is thinking in such a setting without interfering in or altering the observing process.

The literature on interviewing techniques, protocol analysis, and eyewitness testimony all suggest that there are ways and means for an investigator to elicit information from subjects about their thinking. By having subjects think out loud, give retrospective reports, or respond to leading and non-leading questions it is possible to learn about the information that is heeded as subjects observe. The research on those techniques reveals that it is possible to get information that is reflective of the thinking that actually does take place and, furthermore, this can be done without altering the subject's thinking while reporting nor subsequent to the reporting. The following chapter will incorporate such information into a method that uses an interview format to assess the observation ability of grade six students.

# CHAPTER 3: THEORETICAL FRAMEWORK AND METHODOLOGY

This chapter spells out in detail how the findings and procedures discussed in Chapter 2 were incorporated into an interviewing procedure to assess the observation ability of grade six students. The Norris model of observation competence is presented as a definition for the construct and considerable examples are provided of how the interviews with students were conducted. The data recording technique is explained and the procedure for interpreting the data is presented. Finally, evidence of reliability and validity is presented.

## A Conception of Observation Competence

According to Norris (1984), observation competence consists of three broad proficiencies: (a) making observations well, (b) reporting observations well, and (c) correctly assessing the believability of reports of observations. Norris presents sets of conditions that are conducive to making and reporting observations well and a set of principles for assessing the believability of reports of observations. Only the conditions for making and reporting observations are reproduced in Tables 3-1 and 3-2, since the study did not address the activity of assessing observation reports.

Table 3-1

Making Observations Well

In order to observe well an observer should:

1. not allow his or her emotions to interfere with his or her making sound judgements;

2. be alert to the situation and give his or her observation careful consideration;

3. have no conflict of interests;

4. be skilled at observing the sort of thing observed and in the technique being used;

5. have theoretical understanding of the thing observed;

6. have senses and sensory equipment functioning normally;

7. not be influenced by preconceived notions about the outcomes of the observation;

8. use as precise a technique as is appropriate;

9. observe in situations in which good access to the thing observed is available. Access is good to the extent that:

    a. there is a satisfactory medium of observation;
    b. there is sufficient time for observation;
    c. there is more than one opportunity to observe;
    d. if used, instrumentation is adequate. Instrumentation is adequate to the extent that:

        i. it has suitable precision;
        ii. it has a suitable range of application;
        iii. it is of good quality;
        iv. it works in a way that is well understood;
        v. it is in good working condition.

Note. From "Defining Observational Competence" by S.P. Norris, 1984, Science Education, 68, p. 135. 

Table 3-2

Reporting Observations Well

In order to report observations well an observer should:

1. report the observation no more precisely than can be justified by the observation technique that was used;

2. make the report close to the time of observing;

3. report the observation himself or herself;

4. make the report in the same environment in which the observation was made;

5. report the observation in a well-made record, if it is reported in a record. (To make an observation record well an observer should meet the conditions for making any observation report well.)

Note. From "Defining Observational Competence" by S.P. Norris, 1984, Science Education, 68, p. 136. 

Norris's conception of observation competence draws heavily on the work of Ennis (1962, 1980). The specificity of the model is useful because it helps to pinpoint the strengths and weaknesses of students' observation ability and thus supports specific corrective action for instruction. This relates closely to the point raised in the discussion of construct validity in Chapter 2 where it was argued that, in developing tests of mental abilities, it is important to have well defined constructs.

The conditions that Norris says are important for making and reporting observations determine the types of information that are sought through the interviews. While students work through the various activities and make observations, they heed a great deal of information. They choose instrumentation (rulers) with which to measure; they gain access to what they observe; they draw on their theoretical understanding of what they observe; they choose and use techniques to carry out their observations; and, they make records of their observations. These are the sorts of things that, Norris maintains, must be done according to certain conditions if subjects are to observe well. In order to do these things the subjects' minds must be active and draw on pertinent information.

In order to find out why subjects performed as they did, I asked probing questions that were intended to get the subjects to reveal their thoughts. In some cases subjects may have had to search their long-term memory, particularly if asked to give reasons for an action or a response, but much of the reasoning that leads to particular observations is believed to be found in short-term memory.

The Norris model, then, helps to decide what information to seek and what questions to ask. If a subject attempts to gain better access for observing, then questions related to access would be asked. The subject may be asked why a particular vantage point was chosen, how it helps, or if other vantage points would be helpful. When choosing rulers with which to measure, the subject may be asked questions that relate to knowledge of instrumentation such as its working condition, how it is used, or how precise it is for the task at hand. The responses that are given help the interviewer understand what is in the subject's mind, or what is not in the subject's mind, as the observations are made. Through this method a sampling of subjects' proficiency with the various observation conditions is obtained. The various conditions of the Norris model help the investigator decide what information to seek -- in essence, the model guides test construction.

## Method

### Sample

A class of twenty-four grade six students took part in this study. They attended a K-6 school in the Green Bay Integrated School District on Newfoundland's Northeast coast. With one exception, the students had attended only this school during their school years and had all been taught by the same teachers using the same program. The exception had transferred to the school three years previously. The science program at the time of the study was STEM Science (1977), now revised as Addison-Wesley Science (1984). The class represents a range of student ability from those who have experienced considerable difficulty with the school curriculum to those who are very bright academically. The average age of the students was 12.2 years.

For a number of years the issue of female underachievement in science has received attention from researchers. There has been virtual consensus in the literature that, as a group, boys outperform girls even at the youngest ages, although the early differences are not substantial (Erickson & Erickson, 1984, p. 64). The NAEP Report of the 1976-77 Science Assessment (1978) found the male advantage was slightly higher for test items pertaining to the processes of science. No breakdown is offered for individual processes, so as an aside in this study, since the class consisted of approximately equal numbers of boys and girls, a comparison of the performance of boys and girls will be made to determine if the general trend noted in the NAEP Report holds for the process of observation.

Procedure

Data was collected through student interviews conducted over a one week period. First, I met with the students as a group to explain what would happen. They were told that they would be taking part in research that might have an impact on the teaching of science in elementary schools, so it was important that they do their best. Students were told that there would be no pass or failure, and that the results would not be used to assign school grades. They were given the opportunity of not taking part, which two students accepted. However, one of them returned after learning from classmates that the activities were fun and the situation was non-threatening.

Students came one at a time to the room where the activities were to take place. Each student was shown the video camera and given the opportunity to examine and use it for a moment before the session began. They were shown the materials for the activities and instructed to use anything they needed as they carried out each activity. All the activities involved the study of water drops, and the first activity was designed to help them use a dropper to get drops of a consistent size.

Students were instructed that as they did each activity they would be asked questions about what they were doing, what they reported observing, or what they were thinking. They were also encouraged to say anything they wanted about what they were doing or thinking, even when no questions were asked. Each activity was explained to the student before the camera was started. After each activity, the camera would be stopped while the next activity was explained and prepared.

All of the materials that were needed for each activity were laid on a table that served as a working area. For some of the activities the materials were selected by the investigator and handed to the student with instructions on how to proceed. Other materials were selected by the student as required during the progress of an activity. A complete session with a student lasted about 40 - 60 minutes. At the end, students were given a chance to view themselves on the video.

Science Activities

In order to provide a science setting in which students were required to make and report observations, six activities involving the use of water drops and a medicine dropper were chosen. Water drops offer a wide range of opportunities for observation: they can act as magnifiers when they are sufficiently small to retain a curved shape; they can maintain a curved shape or spread out depending on the surface on which they are placed; and, because of cohesive properties, water can be piled on a surface. Various properties such as these were investigated in the series of activities which are described in what follows.

Activity 1: Tiniest drops.

Objective: To use a medicine dropper to produce the tiniest drop of water possible.

Materials: Medicine dropper, wax paper, water.

Procedure: Students use the dropper to produce drops of water. After a few trials they notice that all drops seem to be the same size. By suggestive probes they find that by shaking the dropper they can produce smaller drops, or by putting the dropper very close to the surface they can get very tiny drops.

Activity 2: Magnifying with drops.

Objective: To investigate the magnification effect of placing a drop of water on printed text, and to predict and observe the effect of adding subsequent drops.

Materials: Medicine dropper, water, printed text (preferably gloss paper).

Procedure: Students place one drop of water on one letter of printed text and report what is observed. Normally, they observe that the letter "gets bigger". They are then asked to predict the effect of adding another drop. A second drop is then added to check out the prediction. This procedure is repeated four or five times.

Activity 3: Heads and tails capacity.

Objective: To observe whether the head or tail side of a dime will hold more water.

Materials: Two dimes, water, medicine dropper, tissue paper.

Procedure: Given two dimes, the student will place drops of water on the head side of one and the tail side of the other to determine which, if either, can hold more water. The students are left to their own devices at first to see what procedure they will follow to keep track of the amount of water added. If they prove to be lacking in ability to come up with a procedure, then through a series of leading questions they are guided to counting the drops or matching one-to-one.

Activity 4: Different surfaces.

Objective: To observe differences between water drops on different surfaces.

Materials: Medicine dropper, water, wax paper, tin foil, glass, pencil, note paper.

Procedure: Students place water on the three surfaces -- wax paper, tin foil, and glass -- forming three puddles on each. Puddles are to have one, five, and ten drops each per surface. They are to observe the difference between drops on the same surface and on different surfaces and explain what they have observed. They then make records of what they observed and explain what the records show.

Activity 5: Closest drops.

Objective: To find out on which type of surface two drops of water can be placed closest together without touching.

Materials: Water dropper, tin foil, glass, wax paper.

Procedure: Students place two drops of water as close together as possible, without touching, on the three different surfaces to see if one surface allows closer placement than the others.

Activity 6: Falling drops.

Objective: To observe the effect when a drop of water falls on tin foil from a height of about twelve inches; to measure the lowest height at which this effect is observable; and to determine at what height this same effect is observable when the drop falls on wax paper.

Materials: Water, medicine dropper, tin foil, wax paper, assorted rulers.

Procedure: Students drop several drops of water on the tin foil until the dispersing and contracting effect is noted. Then they will select a ruler with which to measure at what height this effect is first noticeable. They will then attempt to find out at what height the same effect is observable with wax paper.

Probing Techniques Used

As the students carried out the above activities they were probed with questions about what they were thinking and what they reported. The questions were not pre-determined since the activities did not follow an identical course with each student. I had to ask appropriate questions as each activity progressed with each student. When asking questions, two categories were kept in mind: non-leading and leading questions.

Non-leading questions ask subjects to say more about something without providing additional information that would aid the subject in answering. For example, one of the activities required students to place drops of water on two dimes to find out whether the head or tail side could hold more water. As drops were placed on the dimes, a student may have forgotten on which dime the last drop had been placed. Upon the subject's mentioning this, the interviewer might have asked, "Does it matter?". This is a non-leading question because it volunteers no new information to the subject.

If in the previous example the subject had forgotten where the last drop was placed, had mentioned this fact to the investigator, but then proceeded as if it were of no significance, the interviewer might then have asked whether where the last drop was placed mattered. If no response or suitable explanation was offered, the interviewer might then have rephrased it as a leading question: "If you placed two drops of water on one dime before you placed a drop on the other dime, would it matter?". This question offers information to the student and suggests what needs to be considered. Hence, this question is leading.

The basis for selecting leading and non-leading questions as a probing technique relates to the discussion of protocol analysis and eyewitness testimony in the literature review. A conclusion there was that free reports, or responses to non-leading questions, tend to be more accurate than other types of reports although they may contain less information. A second conclusion was that controlled reports, or responses to leading questions, contain more although relatively less accurate information. My aim was to, in the first instance, sacrifice quantity of information for accuracy.

Data

The data consists of what students said and did while carrying out the activities. Video recordings were made of the entire transactions.

Students' talk was categorized as responses to leading and non-leading questions. If a response to a non-leading probe indicated that the student was aware of or operating by a particular competency that Norris has described, then a "+" was tallied opposite that particular competency under a column headed "non-leading". Similarly, if the response indicated the student was unaware of or violating a particular competency, a "-" was tallied. The same procedure was followed with leading probes.

The decision about the competency to which a particular student response is related requires familiarity with the Norris model. Care was taken in deciding the significance of the probe as well as the significance of the response. To illustrate how data was coded and recorded, the following interview was compiled using samples of probes and responses from a number of students. The object of the activity in this interview is to determine whether the head or tail side of a dime can hold more water. The student was given two dimes, a dropper, and some water. Having placed the dimes so that one was showing heads and the other tails, drops of water were placed on one dime until it was covered. The same was done with the other dime. The interview follows; the numbers in parentheses refer to points that are discussed subsequently.

Student: The head side can hold more.

Investigator: How do you know?

Student: Because this one looks like it is piled up more. It's thicker. (1)

Investigator: Is there some way of measuring how much water is on each?

Student: (no response) (2).

Investigator: Can you put more water on each dime?

Student: (Student starts adding more water to each dime. The impact of the falling drop caused the water to shake and spill off one of the dimes.)(pointing) That one there can hold more.

Investigator: Why do you say that?

Student: Because the other one can't hold anymore. It spilled off. (3)

Investigator: Could anything else have made the water spill off?

Student: (no response) (4)

Investigator: Could anything have knocked it off?

Student: Yes, maybe. It may have knocked it off because the water shook when it landed. (5)

Investigator: Place another drop on the other one. Hold it as high as before to see if this one will spill.

Student: (Student does as directed and this water too spills off.) The falling drop knocked it off.

Investigator: O.K. We'll need to start over. How will you know which dime has more water on it?

Student: I'll just put one drop on each dime at a time. (6)

Investigator: How will you know when a dime holds all the water it can?

Student: When it spills over. The one which gets the most without spilling is the winner. (7) (Student started the process again. While placing water on the dimes the student needed to refill the dropper but forgot where the last drop had been placed and therefore where the next drop should go.) I can't remember which dime is supposed to be next. (8)

Investigator: Does it matter?

Student: Yes. (9)

Investigator: What difference does it make?

Student: Because if I put it on the wrong one it will be all fooled up. (10)

Investigator: Why?

Student: Because then I can't be sure which one held the most. (11) (Student discards water from each dime and dries them with a paper towel before starting over.)

Investigator: Was that necessary?

Student: Yes, there was water left on each dime and that might have been enough to make extra drops. (12) (Student continues placing drops on the dimes, occasionally leaning low to peer at dimes from side.)

Investigator: Does that help?

Student: Yes, I can see better how much the water is piling up. (13) (Finally one of the water piles breaks and it spills onto the table). Ah! There. That one holds more water because it hasn't broken yet.

In the paragraphs that follow, I indicate how the interview was interpreted in light of the Norris model. The numbers of the paragraphs refer to the numbers in parentheses in the preceding hypothetical interview. The results of the interpretation are presented in Table 3-3.

1. In response to a non-leading question the student replied with an answer that indicated he or she did not use a precise technique. Therefore, in Table 3-3, a "-" is tallied after condition 8 in column 1 for non-leading probes.

2. In response to a leading question the student was unable to formulate a reply or suggest a means of making an observation. The student seems to be very unfamiliar with the activity at hand and has little understanding of what is to be observed or what technique to use. A "-" is tallied at condition 4 under leading probes.

Table 3-3

Categorization of Responses Received in Sample Interview

| Conditions | Non-Leading Probes | Leading Probes |
|---|---|---|
| 1. Not allow emotions to interfere with his or her making sound judgements; | | |
| 2. Be alert to the situation and give observation careful consideration; | - | |
| 3. Have no conflict of interest; | | |
| 4. Be skilled at observing the sort of thing observed and in the technique being used; | + | - |
| 5. Have theoretical understanding of the thing observed; | - - + + | + |
| 6. Have senses and sensory equipment functioning normally; | | |
| 7. Not be influenced by preconceived notions about the outcomes of the observation; | | |
| 8. Use as precise a technique as is appropriate; | - + | |

(table continues)

Table 3-3 (cont'd)

| Conditions | Non-Leading Probes | Leading Probes |
|---|---|---|
| 9. Observe in situations in which good access to the thing observed is available. Access is good to the extent that: | | |
| a. there is a satisfactory medium of observation; | + | |
| b. there is sufficient time for observation; | | |
| c. there is more than one opportunity to observe; | | |
| d. if used, instrumentation is adequate. Instrumentation is adequate to the extent that: | | |
| i. it has suitable precision; | | |
| ii. it has a suitable range of application; | | |
| iii. it is of good quality; | | |
| iv. it works in a way that is well understood; | | |
| v. it is in good working condition. | | |

3. This response to a non-leading probe indicates a lack of theoretical understanding of what is being observed. Not to realize that the impact of the falling water drop has caused the water to spill seems to indicate that the student doesn't really understand very simple ideas about how drops of water behave. To be unable to bring the required background knowledge to bear in this case is taken as evidence that it is lacking. This is scored as a "-" for condition 5 under non-leading probes.

4. The question asked is non-leading because it offers no information to the student other than a suggestion to think of alternative explanations. Now the student must think about what he or she knows about this sort of thing and apply this knowledge. The lack of response would indicate that the student does not have, or is unable to draw upon, theoretical understanding about the thing observed. A "-" is tallied after condition 5 under non-leading probes.

5. The follow-up question to number 4 is leading due to the addition of the word "knocked". This word carries the connotation that something is being hit. This is rather suggestive and can be expected to direct the student's attention to the possibility of the water being knocked off by the falling drop. The student's response indicated that with a suggestive

probe there is understanding. Here, a "+" is tallied after condition 5 under leading probe.

6. Now, in response to a non-leading question the student demonstrates an understanding which was not displayed in response 1. The student either has learned something through the course of the activity, or maybe is becoming more alert to the situation. In any case, this response is sufficient to warrant a "+" at condition 8 under non-leading probes.

7. This response in reply to a non-leading question indicates the student now seems to have better theoretical understanding than previously indicated. A "+" would be entered under non-leading probes at condition 5.

8. This student report was not in reply to a question. Nevertheless, it does indicate that the student was not alert to the situation and not giving his or her observations careful consideration. A "-" is tallied under non-leading probes at condition 2.

9-11. Even though the student was not alert in the previous example, the student now demonstrates awareness of the consequences of this inattentiveness. In response to three non-leading probes the student demonstrates an understanding of the technique being used. A "+" is entered at condition 4 under non-leading probes.

12. The response this time indicates that the student has very good theoretical understanding of the thing to be observed. It may be possible to suggest that this indicates alertness, skill at observing the type of thing observed, or precision - probably all equally valid interpretations. I choose to say that it demonstrates good theoretical understanding, because all the other suggested possible choices can be subsumed under this one. A "+" is tallied at condition 5 under non-leading probes.

13. This response indicates that the student is aware of the importance of adequate access. Here, the student realizes that shifting position affords a better view of what is being observed, although in this case it doesn't influence the precision of counting drops. A "+" is tallied at 9a under non-leading probes.

When all activities were completed for an individual student, the resulting record sheet showed a number of "+" and "-" tallies beside the various observing conditions as listed by Norris. A composite sheet was also prepared whereby the total of all tallies for all students was compiled onto one record sheet.

Interpretation of Data

The interpretation of data in the following chapter is primarily qualitative. Although there are quantitative aspects to much of the coding, selection, and organization of the data, the data was used to provide qualitative descriptions of student observation competence rather than statistically testable differences among them. For example, suppose that under non-leading probes at condition 8 (use as precise a technique as is appropriate) it is found that there are 4 "+" and 12 "-" tallies. This would indicate that for the most part this student does not seem to consider the importance of appropriately precise techniques for gathering information. This may or may not be what we should expect of grade six students, but it is obviously an area that can be addressed in instruction. The tendency for a student to respond favourably to this condition can be expressed as a decimal. In the previous example the student responded favourably on four out of a possible sixteen occasions. This is .25 of the total possible. By looking at the data in this way it is possible to see general tendencies in the way students behave. Such data can lend support to descriptions such as: the student "tends to" be influenced by preconceived notions; the student "generally lacks" theoretical understanding; the student is "usually" alert to the situation; or the student "shows good competence" in using instrumentation.

The percentage of "+" responses in reply to leading probes gives us less dependable information (Loftus, 1979). The information could serve to indicate the presence of knowledge with the competencies, but because it took a leading probe to prompt the student to draw on that knowledge, then the student didn't see on his or her own the relevance of that knowledge to the situation at hand. The leading probe may prompt the student to understand the significance of certain conditions or the relevance it has to what is being observed. The failure to apply their competence could be a result of not understanding that the general knowledge they possess is transferable and applicable in many situations, not considering alternatives, or not being imaginative. In any case, the percentage of responses that can be rated as "+" gives reason for using such descriptive phrases as "ready to learn" or "tends to possess (or lack)". For example, although subjectivity is involved, there would seem to be good reason to say that a student is "ready to learn" about instrumentation if he or she displays competence in 40-60% of the opportunities to do so in response to leading questions. Similarly, if a student displays understanding in only 10% of the opportunities provided, it would seem like a fair judgement to say that the student "is lacking" in theoretical understanding. Students may be found able to consider with little prompting various factors, such as, background knowledge, and need for being alert and precise, as observations are made. On the other hand, they may be so weak in knowledge of the competencies that they are observationally inept. Such information can be useful in deciding what level of instruction might be used with such students.

Reliability and Validity

An "inter-rater" reliability, using the same judge to re-score the tests, was calculated. Ten of the subjects' interviews were rescored by the same judge 18 months later and the two sets of scores correlated. The results are presented in the following chapter along with the rest of the data.

Criterion-related evidence for validity relies on the extent to which some outside variable correlates with the construct being measured. For example, given that the science processes--including observation--are typically taught to students in current school programs, one might expect a correlation between school science grades and performance on a test of observation. Such a criterion-related study was conducted and the findings are presented in the results section.

Content-related evidence for validity rests on the extent to which the method for observing the students covered the construct of interest. The Norris model of observation competence, as presented earlier in this chapter, provides the most extensive set of criteria for judging observation competence. The model includes an extensive list of conditions which, if met, help to ensure that an observer has the ability to make accurate observations.

There is a limitation, however, with the extent to which the model is covered by the interviews. Certain conditions of being a good observer could not be assessed because the situation was such that the conditions were met necessarily due to the context in which the study took place. For example, in school science one does not typically find conflicts of interest or emotionally charged activities. Nor was there any indication that students in this study suffered from improperly working senses. Conditions such as these were not assessed in the study. However, a sufficiently broad sampling of the conditions was assessed so as to allow valid descriptions of students' observation ability.

In addition to coverage of the construct, there must also be evidence to show that the types of items are suitable to the age group under consideration. The activities were selected on the basis of being representative of types of activities that can be found in sixth grade science programs. They require students to investigate phenomena through use of skills and procedures that they presumably have developed after a few years of schooling. Skills such as measuring with rulers, counting, making comparisons, and drawing sketches were to be employed in carrying out the activities.

A search of the students' previous texts revealed that the activities were not included in the current program. The likelihood of the students having previously performed the activities was further reduced by the fact that they were not taken from any published source. Instead, they were contrived by the writer especially for the study. Consequently, the chances of students relying on simple recall to make and report their observations were lessened. Care was taken to ensure that the activities were not too difficult or too easy. It was important that the activities not require subject knowledge that was beyond the capability of normal grade six students to comprehend. Similarly, the activities would have been unsuitable if they were overly simple and dealt with material that was far below the level of grade six. The judgement of the investigator along with that of an expert in science education led to the conclusion that the difficulty level of the activities was appropriate.

In addition to the expert judgements of suitability of the activities, a trial was conducted on four grade six students. At the time of the trial, consideration was being given to using a series of activities involving water drops, pendulums, pulleys, inclined planes, and static electricity. But, in working with the trial group, it was found that the amount of materials required, the amount of time required to set up materials, and the amount of explanation required for each activity resulted in unnecessary complications and inefficient progression through the testing situation. A decision was then made to go only with the water drop activities.

Also, care was exercised to ask questions that would prompt students to reveal information about their observation competence. The aspects of the Norris Model and of the theory of protocol analysis guided the formulation of questions. Questions were phrased carefully so that the degree of leadingness was always controlled and tempered to the course of the activities and the performance of the student. The extent to which these concerns were addressed during test development is further evidence of validity.

In summary, it seems reasonable to claim that the procedure that has been devised produces valid descriptions of a number of aspects of grade six students' observation competence as they work through a series of science activities that involves the use of water drops. The activities encourage students to draw on their observation skills, and the questions asked are suitable. This description of the methodology clears the way for the presentation and interpretation of data which follow.

# CHAPTER 4: RESULTS AND DISCUSSION

The first portion of this chapter presents the quantitative data that was derived from the analysis of student interviews. It shows how the group of students performed overall with the various aspects of observing and provides the averages which are used to depict the typical student. General trends are discussed and overall strengths or weaknesses are noted.

The second portion of the chapter presents the qualitative descriptions that form the core of the study. The typical student is described and in addition, an average, a below-average, and an above-average observer are described in detail according to the things they did and reported during the interviews. Their specific strengths and weaknesses are identified. This chapter answers the question, "How competent are grade six students at observing?".

## Reliability Results

Eighteen months after the initial coding, five boys and five girls were selected randomly from the total sample. Their video tapes were recoded by the same judge and correlated with the initial coding. The obtained correlation coefficient was .89, which by current standards is quite acceptable to support the claim that the testing procedure is a reliable one.

## Criterion-Related Validity Evidence

A criterion-related validity study was conducted to determine the relationship between observation scores as obtained in this study and school science grades. Logically, if observation skill is taught in the school science program, then student achievement with the process might be expected to be in line with achievement in science generally. However, the obtained correlation figure of .34 indicates that this is not so.

When we look at what the school science grade reflects, we begin to see that there is really no reason that there should be a high correlation. School science involves much more than observing. A great deal of time is spent with content, conceptual understandings, and vocabulary. Other science processes in addition to observing are practiced. Notwithstanding the importance of observation as a fundamental skill, it is very likely that students' performance in school science, as reflected by their science grades, is not tied to their observation ability. This is further reinforced by the fact that the conditions for observing well, as described by Norris and reflected in the students' observation scores in this study, are not specifically taught to students. Students may be given opportunities to observe without necessarily being taught how to observe well. The low correlation, in this instance, may indeed be evidence to support the notion that the conditions for observing well have not been taught to students.

## Quantitative Descriptions of Student Observers

Three conditions of making observations well and three for reporting observations well were dropped from the analysis due to insufficient data or due to irrelevancy in the content. The conditions for making observations well that were dropped are: (a) not allow emotions to interfere with making sound judgements; (b) have no conflict of interest; and (c) have senses and sensory equipment functioning normally. The three conditions for reporting observations well that were dropped are: (a) make the report close to the time of observing; (b) report the observation himself or herself; and (c) make the report in the same environment in which the observation was made.

These six conditions are important for good observation. However, it did not make sense to assess some of them in the context of this study, because, for instance, the situation demanded that the three conditions for reporting observations had to be satisfied, and, for others, not enough data was collected to justify any conclusions.

An examination of condition 9 in Table 3-1 shows four subgroups, one of which is further subdivided into five subgroups. Within the individual subgroups very little data was collected. Therefore, the groupings were collapsed to form two conditions: conditions 9a, 9b and 9c were combined and conditions 9di, 9dii, 9diii, 9div, and 9dv were combined. Data is thus reported on the 9 conditions listed in Table 4-1.

Table 4-2 presents the response rate for all students to the 9 conditions in Table 4-1. The first column of numbers gives the total number of tallies that were recorded for each condition. The second column indicates what portion of the responses in column 1 was considered positive or indicated competence. Columns 3 and 4 give a breakdown of columns 1 and 2 for boys. Similarly, columns 5 and 6 break down the same data for girls. The row at the bottom of the table, labelled "composite", represents the data obtained from the totals of all the individual conditions.

Note that there were many more non-leading probes than there were leading probes. This is a result of the tendency to ask first a non-leading question, and, depending on the response, it may have been followed with another non-leading question or perhaps a leading one. Inferences that rely on the leading probe data are less well founded than inferences based on non-leading probe data.

Some cells have little or no data. This is true for leading probes in the preconceived notions and precision of report conditions where no data was available. There were very few leading probes for the instrumentation condition.

Table 4-1

<u>Conditions for which Data is Reported</u>

1. Be alert to the situation and give his or her observation careful consideration.

2. Be skilled at observing the thing observed and in the technique being used.

3. Have theoretical understanding of the thing observed.

4. Not be influenced by preconceived notions about the outcomes of the observations.

5. Use as precise a technique as is appropriate.

6. Observe in situations in which good access is enhanced by a satisfactory medium, sufficient time, and more than one opportunity to observe.

7. Use instrumentation that is adequate due to suitable range of application, suitable precision, good quality, good working condition, and understood way of working.

8. Report the observation no more precisely than can be justified by the observation technique that was used.

9. Report the observation in a well-made record if it is reported in a record.

Table 4-2

Total Number of Responses Per Condition and Rate of Positive Response

| Conditions | | Whole Group TR[a] | Whole Group PRR[b] | Boys TR | Boys PRR | Girls TR | Girls PRR |
|---|---|---|---|---|---|---|---|
| Alertness | NL[c] | 102 | .43 | 41 | .41 | 61 | .44 |
| | L[d] | 73 | .62 | 30 | .63 | 43 | .60 |
| Technique | NL | 61 | .32 | 25 | .60 | 36 | .28 |
| | L | 31 | .71 | 13 | .85 | 18 | .61 |
| Theoretical Understanding | NL | 86 | .41 | 38 | .42 | 48 | .42 |
| | L | 56 | .43 | 23 | .52 | 33 | .36 |
| Preconceived Notions | NL | 24 | .37 | 12 | .33 | 12 | .42 |
| | L | 0 | - | 0 | - | 0 | - |
| Precision of Technique | NL | 63 | .35 | 30 | .37 | 33 | .33 |
| | L | 34 | .68 | 14 | .64 | 20 | .70 |
| Access | NL | 76 | .41 | 32 | .44 | 44 | .36 |
| | L | 29 | .76 | 12 | .92 | 17 | .65 |
| Instrumentation | NL | 73 | .62 | 24 | .75 | 49 | .55 |
| | L | 16 | .88 | 3 | .33 | 13 | 1 |
| Precision of Report | NL | 36 | .53 | 17 | .41 | 19 | .74 |
| | L | 1 | - | 1 | - | 0 | - |
| Well-Made Record | NL | 29 | .28 | 12 | .25 | 17 | .29 |
| | L | 29 | .72 | 13 | .77 | 16 | .69 |
| Composite | NL | 561 | .41 | 235 | .42 | 326 | .41 |
| | L | 272 | .64 | 112 | .69 | 160 | .61 |
| | NL + L | 833 | .49 | 347 | .51 | 486 | .48 |

Note. [a]Total number of responses [b]Positive response rate = ratio of positive responses to total number of responses [c]Non-leading [d]Leading

<u>Instrumentation</u>

Students demonstrated considerably better performance with instrumentation than they did with any of the other conditions. When one considers the types of instruments used in the study, then a possible explanation for this performance can be offered.

The instruments used in this study were rulers. Students are exposed to rulers almost on a daily basis from the time they begin school. They get to use the instrument in many situations and become proficient in their use by the time they reach sixth grade. I would speculate that had the study required more varied instrumentation such as balances, thermometers, or capacity measuring devices, then the demonstrated level of competency would have been somewhat lower for this condition.

<u>Leading Probe Effect</u>

The positive response rate columns indicate a strong leading probe effect. Students generally displayed more competence in response to leading questions than they did to non-leading questions. Two exceptions to this effect occurred for the theoretical understanding condition with girls and for the instrumentation condition with boys. In the latter case the variation is unreliable because there were only three responses altogether. However, for the former there is no apparent reason for the variation.

It is to be expected that leading probes will elicit more favourable responses than non-leading probes because they carry information that can assist the students in formulating a response or deciding what to do. The important point for consideration is the significance of this trend for the observation ability of students.

The leading probe effect possibly can be explained by reference to the notion of dispositions as discussed in the critical thinking literature. Students may have the skill to be good critical thinkers but not be disposed to use that skill. It is possible that this happened with students in this study. They have the knowledge and ability to perform better, but do not search well for the relevant information. The stimulation of leading questions sends them searching in the right direction and their ability to make and report observations is significantly improved.

Note in Table 4-2 the two extremes of leading probe effect. The well-made record condition shows the greatest difference in positive response rate for non-leading and leading probes: .28 and .72, respectively. The positive response rate of .28 is also the lowest for non-leading questions. Students have a low understanding of the need and the means of making records of their observations, and demonstrated very poorly created records in the form of sketches, diagrams, and charts. However, given leading questions their competence rate jumps. Such a vast improvement in response to leading questions gives rise to some speculation.

Students at this age are typically at the threshold of abstract thought. Making records can essentially be thought of as an abstract activity, requiring students to transfer images of what was observed to a more symbolic paper-and-pencil form. Their low performance in response to non-leading probes indicates their lack of knowledge with this condition, however, their much more typical performance in response to leading questions indicates their readiness to learn this competency and to deal with abstract activity.

Another line of speculation relates to program or instruction. It is possible that poor programming or inadequacies in instruction has meant that students have not had many opportunities to develop this competency. They may have performed numerous activities without having been required to make records of their observations. The effect on students has been that they are inefficient in producing records, however, they can very quickly learn when simple records, like the ones in this study, are required.

Student competence in the theoretical understanding category in response to leading questions was only .02 better than for non-leading probes. Although students demonstrated average performance with this condition, when compared to the composite for non-leading probes, there was no leading probe effect. A line of speculation can be offered which may account for the lack of leading probe effect.

It may be that subjects have very limited theoretical understanding that they can draw on. Given the fact that the study involved twelve-year-olds, it is reasonable to expect that subjects' theoretical understandings of the world in general, as well as of specific concepts and ideas, have not been clearly and fully developed. Hence they were able to avail of their limited theoretical understanding to the extent that they did in response to non-leading questions but were not equipped to fare much better in response to leading questions. Since leading questions were intended to get subjects to see the relevance of already held information, not to instruct, then no leading probe effect can be expected if the required information is just not present.

Sex Effect

One final point to be noted in Table 4-2 is the comparison between boys' and girls' scores. It is immediately apparent that there is no appreciable difference in the performance of boys and girls. There are naturally some variations from the average but this is to be expected. However, the variations are not consistently in favour of either sex and the differences are generally less than .1 with the overall composite being only .03 in the difference for the two sexes. During the course of the interviews there was no impression that one sex was outperforming the other and the data verifies this. Variations then can be attributed to individual differences, not sex differences. This result is at odds with the NAEP (1978) results, mentioned under the section describing sample, which reported that boys perform slightly better than girls on the science processes.

## Qualitative Descriptions of Student Observers

The qualitative descriptions are presented in the following order. First, there is a description of the typical grade six student as determined from the whole group positive response rates as presented in Table 4-2. Following the typical student description come descriptions of 3 actual students: an average observer, an above-average observer, and a below-average observer.

### The Typical Grade Six Student

The typical grade six student is not very proficient at making and reporting observations. Overall this student demonstrates competence in 41% of the opportunities provided when non-leading questions are asked. When leading questions are asked the student demonstrates competence in 64% of the opportunities to do so. So, while the second figure might give an indication of the student's actual capability, the first figure gives information about how the student performs independently.

Alertness. The typical grade six student tends not to be very alert and does not seem to appreciate the importance of being alert. In fewer than half of the opportunities to indicate proficiency with this condition did the student respond in a way that indicated competency. For example, it is quite common for the student not to notice that the ridges and crinkles on tin foil might interfere with the drops of water being placed there. It is quite routine for the student not to notice two or three drops of water coming out of the water dropper when the activity at hand requires the drops to be counted. As a result, it is very common for the student to overlook factors that interfere with the accuracy of observations and to fail to realize that certain variables must be controlled. Often, the student tends to behave and to perform activities in a nonchalant manner. Even when probed with leading questions, the student, although showing more indication of proficiency than with non-leading probes, shows competence just over 60% of the time.

Technique. The student shows very limited skill in observing the things observed and in the technique being used. In only 32% of the cases is proficiency indicated. For the most part it can be said that the student is not very skilled at observing water drops, and, if one can generalize, is not very skilled at observing simple, everyday phenomena. The student appears to struggle with the techniques to be used in structured and controlled situations and is very hesitant and uncertain as the activity is carried out. This is evident in the Heads and Tails Activity that requires testing dimes to see which side can hold more water. Almost invariably the student does not think of the necessity of measuring the water in some way, and, when led to realize that drops can serve as a measuring unit, the student still has trouble proceeding until led to understand that the drops must be counted or else placed alternately on each coin. While proceeding with the technique, the student is prone to many errors such as miscounting, forgetting where the next drop is to be added, or holding the dropper too high thus causing a spill on impact. There is a general lack of care and self-confidence and the student needs to be watched, guided, or advised in order to avoid errors, oversights, and miscalculations in the technique. It is promising, however, to note the difference when one considers the competence indicated in response to leading questions. In 71% of these cases the student indicates proficiency, as if the competence lies there, latently, just waiting for the appropriate stimulation.

Theoretical understanding. The typical student does not possess a great deal of theoretical understanding of the things observed. There appears to be a general lack of understanding of simple everyday phenomena. In only 41% of cases is the student able to indicate theoretical understanding of what is observed. This theoretical understanding need not be very profound or elaborate, and here refers to just common knowledge of everyday things and events. For example, as previously mentioned, the student seems not to understand that the impact of a falling drop could cause the water to spill off the dimes, or that the ridges or indentations on tin foil can affect the shape, outline, or even position of water drops placed on it. These rather obvious understandings become apparent to the student when leading questions are asked. However, other conceptual ideas, such as that water adheres to wax paper more than to tin foil or that water drops do not spread and contract when they fall on wax paper, are more foreign to the student's knowledge base. Even with leading questions, the student fared little better and was unable to demonstrate much understanding about the phenomena involved with water drop activities. Thus, the theoretical understanding seems not to be there just beneath the surface, waiting to be tapped. It just seems not to have been acquired to any degree.

<u>Preconceived notions</u>. The student appears to be influenced by preconceived notions and tends to report observing what was expected, even if quite the reverse occurred. For example, in the Magnifying with Drops Activity, the student can be expected to report that a second drop of water added to a first drop will produce more magnification than the single drop. This error of

observation was demonstrated repeatedly; it took normally three or four attempts at adding water before the student would report accurately. Similarly, the student would tend to report a spreading and contracting effect of a water drop falling on wax paper, after having observed the effect on tin foil and being asked to find the height at which the dropper must be held to observe the effect on wax paper. But rather than report that it doesn't seem to happen on wax paper, the student tended to report observing it happen and provided a measurement of the height from which the effect was created.

Precision of technique. The student is not inclined to choose a precise technique or to assess a technique to determine if it is precise enough. If the observations require some means of quantification, the student is prone to overlook this and merely use a visual check to compare amounts. Even when the student realizes that a more precise technique is required, it is not unusual that unorthodox means are used. For example, instead of using a ruler with standard units, the student might use hand spans. When measuring the height of something, the student might just place a finger in mid-air at some seemingly corresponding height and then measure the distance from the spot marked in the air to the floor. This might be done with a 30 centimetre ruler which has to be moved three or four times in order to cover the height in question. The lack of skill with this competency results in the student consistently making inaccurate or imprecise observations.

Access. Good access to the thing observed is another condition whose importance the grade six student doesn't appear to value or understand. For example, the student is quite content to make observations from less than an ideal vantage point as in the Different Surfaces Activity. When observing and sketching drops of water on different surfaces there is a tendency to draw them from what can be observed looking obliquely at the surface. The student doesn't seem to think to look at them from differing perspectives, such as from above or from a side view. Also, there is a tendency to be hasty with observations such as when sketching the drops just mentioned. The student doesn't take sufficient time to have a close look at the objects in question and to see the relative sizes. Instead, the tendency is towards making a cursory type inspection. It is also generally not normal for the student to make repeated observations of a thing, but instead is quite content to rely on a single observation. This frequently is exhibited by the student when investigating the magnifying effect of a drop of water on print. When the student placed a second drop on top of the first and reported that the image had gotten bigger or smaller the student reported being certain about this observation and didn't recheck on another unit of print. The student typically asserted confidence in the first observation.

Instrumentation. Notwithstanding the comments under "Precision of technique", it is with instrumentation that the student displays the greatest degree of competence. The student shows reasonably good understanding of the way instruments are to be used and realizes that they have a limited range of application. For example, the student experienced little difficulty in measuring distances using a 30 centimetre ruler, and in understanding the purpose of rulers that have a greater range (such as a metre stick). The student typically selects the proper range ruler. However, the student sometimes uses a short ruler when a longer one would be more appropriate.

A weakness is that the student doesn't give due consideration to the instrument being in good working condition. For example, it is not uncommon to have the student use a metre stick with a small section broken off and make measurements with it without regard to the missing part.

Generally, when the student is to make a straightforward measurement, then the student is capable of doing so. There may be slight errors due to the ruler not being properly lined up (alertness) or in setting beginning and ending points, but these are basically not problems for the typical student.

Precision of report. When reporting observations the student has a tendency to report them more precisely than can be justified by the technique that was used. For example, after measuring a height by repeatedly moving a thirty centimetre ruler upward through the air until it reached a point in mid-air, the student might calmly report that the height was 94.5 centimetres. Or the student may have attempted to decide whether the head or tail side of a dime could hold more water simply by visual inspection of the total amount of water and keeping no track of the water that goes on the dimes. For example, the student would report that one side holds more than the other by observing how high the water has heaped up. However, this cannot be a precise report for the technique itself is faulty. In general, the student doesn't seem to realize that precise reports require precise techniques.

Well-made record. Probably the weakest aspect of all with regards to the student's observation ability is in reporting the observations in a well-made record. There seems to be little awareness that records are later reviewed, often by a person who has not had the benefit of making the observations in question. The student is alert only to the immediateness of the situation and is unlikely to include the details needed for a complete and thorough record. For example, in the Different Surfaces Activity while preparing a record of drops of water on tin foil, wax paper, and glass, as shown in Figures 4-1a, 4-1b, and 4-1c,

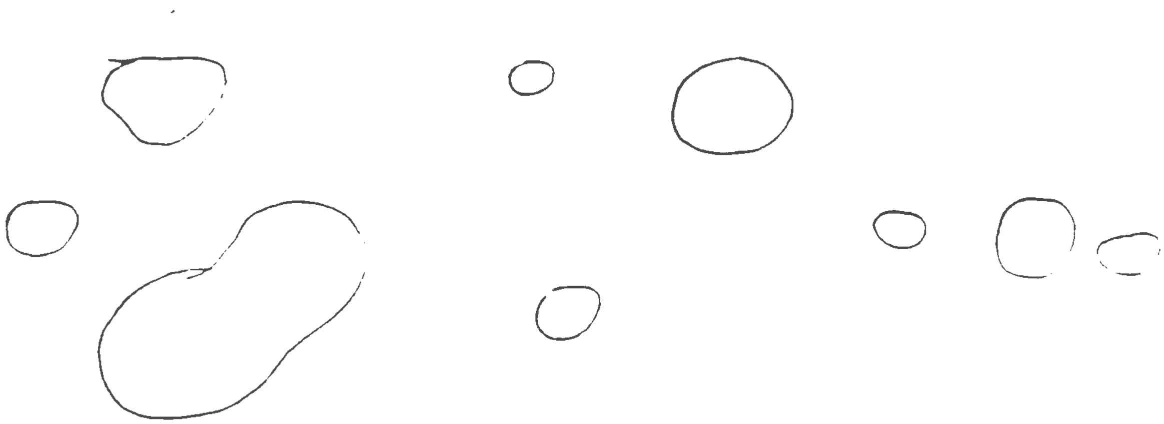

Figure 4-1a. Typical Student's Record - No Probing

Figure 4-1b. Typical Student's Record - Additional Perspective Following Leading Probes


Figure 4-1c. Typical Student's Record - Supporting Information Following Leading Probes

the typical first behaviour is to draw a sketch of the drops from above. Typically, the sketches would be haphazardly arranged on the page as shown, and quite often lack the proper proportions relative to each other. Generally, there is no clarifying information and drops are drawn from only one perspective. After being led to realize that another perspective, for example, a side view, will offer additional information about the nature of the drops, the student's record may be improved to look like the one in Figure 4-1b, albeit still lacking the information which clarifies why the drops are different. Finally, after being led to realize that information needs to be written in, the records may look like that in Figure 4-1c. However, for the most part, the investigator had to provide leading probes to get the student to go beyond the initial sketches which showed only how the drops were spread out. It would appear that this typical grade six student could benefit from practice with making records of observations.

Tammy -- An Average Observer

Compared to the performance of the group as a whole, Tammy is an average observer and possesses very similar traits as the typical student previously described.

Alertness. Tammy is less than adequately alert when making observations. During the Different Surfaces Activity she was making heaps of water containing one, five, and ten drops from the water dropper. As she was adding drops to one of the larger heaps, she stopped, looked up at the investigator and asked, "How many was that?" Twice during the activity her lack of attention caused her to lose track of the number of drops she had placed.

At another time, during the Heads and Tails Capacity Activity, she was placing drops alternately on each dime rather than counting how many drops were being placed on each. At one point she paused for some reason and then, when she was ready to resume, she stated, "I'm not sure where this next drop goes."

"Does it matter?" she was asked.

"Yeah, if I put it on the wrong one it won't be fair," she replied.

Even though she understood the consequences of not being alert, it seems that the slightest distraction such as having to refill the dropper can interfere with her concentration and cause her to forget where the next drop should go. She has trouble directing her attention at one phenomenon for an extended time and doesn't seem to keep in mind that inattentiveness will have detrimental consequences.

On another occasion she was able to show that in response to leading questions she can become more alert. This was shown in the Falling Drops Activity where she was seeking to observe the spreading and contracting effect of a drop of water falling on tin foil. The first time she let a drop fall she was asked, "What happened?"

"Spread out," she answered.

"Did it stay spread out?"

"Yeah," she replied.

She was told to try a few more drops to be sure. After a few tries she was asked again, "Are you sure it stays spread out?"

She answered, "It goes out and comes back again." Even though this effect is very obvious to see, it had taken several attempts before she finally noticed that the water drop did not remain spread. However, it was the suggestive probe that made her realize there must be more to see.

Overall, her tendency is to be rather inattentive to the course of the activity and to overlook obvious things that are important.

Technique. Tammy's skill in observing the sort of thing observed and in the technique being used is also low. A lack of experience seems to greatly inhibit her ability to learn quickly a specified observation technique or to select a technique to use. For example, during the Heads and Tails Capacity Activity, she jumped to a conclusion having added only two or three drops to each side.

She was asked, as the activity began, "How will you go about it?"

She responded, "What do you mean?"

"How are you going to find out which side can hold more water?", she was asked.

She didn't reply, but after just a few drops were added, she stated, "This side."

"How do you know?"

"This side is bigger," she answered referring to one of the dimes which she believed had more water on it.

After being encouraged to start over and keep track of how much water goes on the dimes she was asked, "How will you tell which holds more?"

"I was putting one drop on each," she responded.

"How will you know then if one holds more than the other?

"When it's all filled in," she replied, indicating that she still had not fully understood the technique.

She was encouraged to keep adding water to the dimes and finally, one of the heaps spilled over. "Do you know which one can hold more water?" she was asked.

"This one," was the reply (pointing to the one that had not spilled).

"How do you know?"

"Because when I put the last drop on this one, it spilled," she stated.

"Have you put a last drop on the other one to see if it will spill?"

"No."

She then added a final drop to the other dime and was asked, "How can you tell if the tail has more?"

"Because it's still high. Didn't leak over," she replied.

Even though she had finally grasped some understanding of the technique and gave reasonable answers to the questions to indicate some degree of competence, she had really been led through the activity by the series of suggestive probes.

Theoretical understanding. Tammy seems to lack a great deal of common sense knowledge about the world in general. In the Different Surfaces Activity, for example, she was oblivious to the crinkles and indentations of the foil.

The investigator pointed to some rough and smooth places on the foil, and asked, "Does it matter where you put the drops? Any difference if you put it here, here, or here?"

"No," she replied.

"Does it matter if it's level or not?"

"No," she replied again, offering no further comment.

In the Heads and Tails Capacity Activity, she at one point made the water spill off because she had held the dropper higher than normal.

She was asked, "What happened to the tail side that time?"

"It leaked over," she said.

"Why did it leak over?"

"'Cause it was too much water on it," she replied.

"Could something have made the water spill over?"

"Just too much water on it," she reaffirmed.

It was then pointed out to her that she had held the dropper higher and the drop had fallen further. "Could that make it overflow?" she was asked.

"Yeah," she replied.

"Could you have gotten more water on then?" she was asked.

"Maybe," she answered.

However, after going through this series of questions she was finally asked, "What might happen if the dropper is held too high?"

"Gets too much water on it and overflows," she responded. Even though she seemingly had come to realize the effect of holding the dropper too high, she now had demonstrated that indeed she does not have theoretical understanding of the simple phenomenon involved. Her lack of knowledge about what she was observing was a serious detriment to her ability to make observations.

Preconceived notions. There was very little data that relates to Tammy's being influenced by preconceived notions. However, in one instance it appeared that she may have been unduly influenced by her expectations. This occurred in the Magnifying with Drops Activity.

After she had placed a drop of water on print, she was asked, "Did you see anything happen?"

"Got bigger," she replied.

"What do you think will happen if you add another drop?"

"I'm not sure," she answered.

"What should you expect?" she was asked.

"Get bigger," she responded.

"What happened?" she was asked after adding a drop.

"It got bigger," she said.

This continued for three more tries and then she said, "It stayed the same."

"So, how big is it now?" she was asked.

"The same size as the other letters," she replied.

It appears that Tammy is strongly influenced by preconceived notions here. She maintained that the letter had gotten bigger and bigger, and suddenly declared that it was the same size as the other letters. However, this effect should have been observed after only the second, or at most third, drop had been added. No further instances occurred with Tammy that related to this proficiency.

Precision of technique. Using as precise a technique as is appropriate is an area where Tammy shows a little more competence than the typical student, although she displays competence in only about half of the occasions to do so. It is not unusual for her to choose a suitable technique to carry out observations but then moments later use the technique in a most imprecise manner. For example, in the

Falling Drops Activity, she decided to use a ruler, which under the circumstances was a precise enough method to use. She laid down the dropper, selected a ruler, again held the dropper in mid-air where she believed it had been initially, and proceeded to make her measurement.

"At what height is it?" she was asked.

"About 4 or 5," she responded.

When she was asked, "Are you sure you replaced the dropper at the correct height?" she didn't reply, but instead repeated the procedure of letting the drops fall, but this time she held the ruler beside the dropper as she proceeded.

During the Heads and Tails Capacity Activity, she realized, with some suggestive probes, that she would need to keep track of the amounts of water on the dimes. When asked, "How will you tell which holds more water?" she replied, "I was putting one drop on each," meaning that for each drop she placed on one dime she was placing a corresponding drop on the other. This technique was precise enough but, as was indicated in the previous discussion, she had arrived at this technique through the leading questions posed by the investigator.

When asked in the Closest Drops Activity how she was going to tell which was closest, she replied, "Just look down on the drops."

She was asked if a ruler would be helpful and she said, "No. They're too close." The technique she had chosen in this case was appropriate because close visual inspection was sufficient to compare the distances at which the drops were spaced from each other.

It would seem that her competence in this area is best illustrated in response to leading probes because, as can be seen from the first two examples, her responses to non-leading probes had not been favourable.

<u>Access</u>. Tammy does seem to have reasonable competence with gaining good access to the thing observed. She doesn't seem to be in a hurry to carry out her observations and for the most part is willing to make repeated observations if not satisfied with her first attempts.

For example, after just one attempt in the Tiniest Drops Activity she had looked up and asked, "Can I try again?"

When asked why, she said, "I might be able to make a smaller one." This desire to make repeated observations was illustrated on different occasions as, for example, when she was testing the effect of a drop falling on tin foil or making repeated attempts at the same height to see if the effect could be noted on wax paper.

There were times, however, when she didn't obtain the best vantage point for observing. For instance, in preparing to sketch the shape of drops in the Different

Surfaces Activity, from her normal seating position she looked at the drops and then attempted to draw them. She didn't lean over the drops to view them from above, nor did she lean down to get a side view. She did not try to view the drops from different angles to see if they afforded better or more useful vantage points.

When she had made her sketches she was asked, "What differences do you show?"

She replied brokenly, "Spread out . . . on glass. Wax . . . don't run."

"How about the tin foil?"

"Well that one there is in one place but that one spreads more."

She was then asked, "Is there another way to look at the drops?"

"Sideways," she replied.

"What difference would that make?"

"See how high they are," she answered.

Now, after being led to try a different perspective, she peered at the drops from the side view and proceeded to make sketches. However, she still didn't position herself as well as she should have. She tilted her head and slid down a little in her seat, but she did not get down to the level of the drops.

Although Tammy does take the time to make her observations and is willing to make repeated observations, her lack of effort in securing a good vantage point can be a serious drawback, because, no matter how many observations are made and no matter how long is taken with making them, if they are limited by a less than satisfactory vantage point they are questionable.

<u>Instrumentation</u>. Tammy has a reasonably well developed conception of instrumentation. She understands that some instruments are designed with more precision than others. In one instance she was selecting a ruler to make some measurements. She picked up a centimetre ruler that was marked in coloured centimetre blocks. No numbers were on it so anyone using the ruler would have to count the coloured blocks to obtain a measurement. She looked at the ruler then put it back and selected a normal 30 centimetre ruler with the usual numbering system on it.

"What was wrong with the other ruler?" she was asked.

"There's no centimetres on it," she replied. "And no halves and that."

She understands such simple things as that a metre stick has a greater range of application than the ordinary thirty centimetre ruler, and was able to choose the appropriate ruler when needed to make measurements. She understands how such instruments work and showed good facility with reading the required information from it.

A surprising and conflicting factor was her failure to realize the importance of the instrument being in good working condition. For example, in one situation she used a broken metre stick that had fifteen centimetres missing from the beginning end. Not only did she not account for the missing fifteen centimetres, but she didn't even notice that they were missing. This can probably be related to her lack of alertness.

Precision of report. In reporting her observations Tammy tends not to report them any more precisely than can be justified by the technique being used. For example, it was mentioned that Tammy was measuring the height of the dropper in the Falling Drops Activity.

She said, "At about 4 or 5."

The important word here is "about", because it is rather difficult to discern a cutoff point for the phenomenon, which is a spreading and contracting effect when the drop hits the tin foil. The effect becomes less and less visible as the height diminishes but it is difficult to say exactly when it starts and stops. Her use of the qualifier "about" indicates that she realizes there is a margin of error involved. This is in contrast to several students who reported such a measurement to the half and even quarter centimetre.

For the most part, Tammy made her observation reports accurately based on the method used. However, very few opportunities were afforded to test this competency for it lends itself best to reports of measurements and there was only one activity that required the direct use of a measuring instrument.

Well-made record. When Tammy reports her observations in a record she does try to display a well-made one, but is rather insensitive to the need for thoroughness in the report. While drawing reasonably well organized and laid out sketches of drops of water on different surfaces, she was not immediately aware of the necessity to label her diagrams or to display alternative perspectives so that they would be more meaningful for later interpretation and analysis.

For example, when she drew the sketches for the Different Surfaces Activity, they looked like Figure 4-2a. She included no identifying information.

She was then asked, "What differences are you showing here?"

She pointed at her sketch and said, "All spread out on glass. Wax . . . don't run."

"How about on the tin foil?"

"Well, that one there is in one place but that one there spreads more," she replied.

"Is there another way to look at the drops?" she was asked.

"Sideways," she quickly replied.

"What difference would that make?"

"See how high they are," she responded.

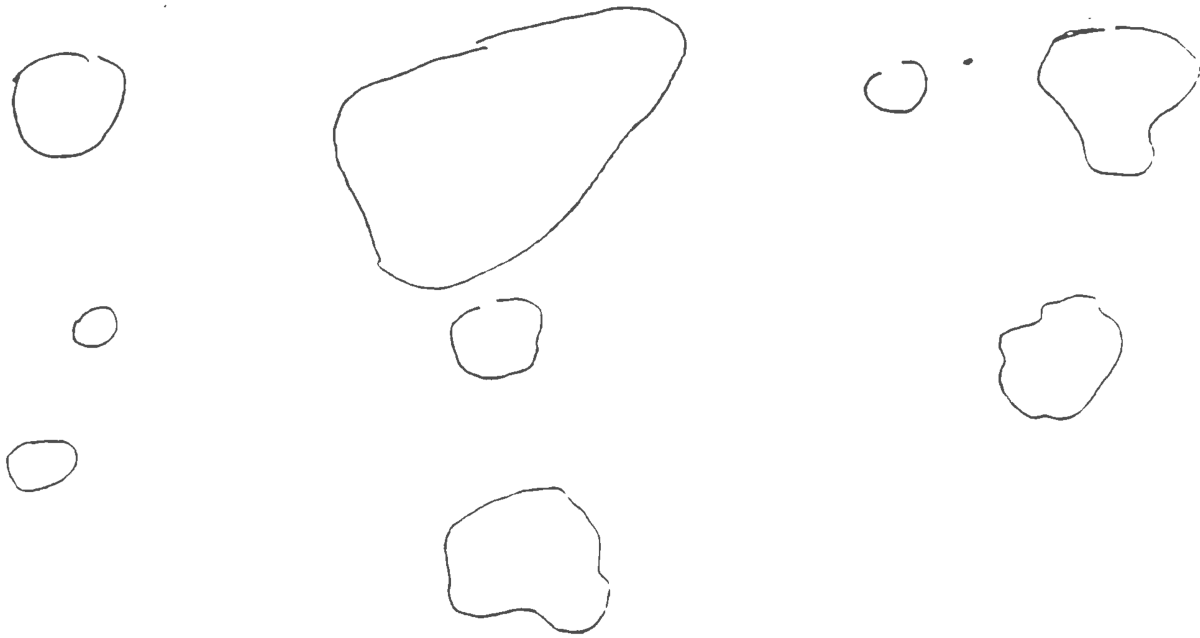

Figure 4-2a. Tammy's Record - No Probing

She then attempted to draw the drops from a side view but she didn't know how to go about it. The investigator had to do an illustration for her and then she proceeded to draw the sketches from a side view. By now her record had grown to look like Figure 4-2b.

She was then asked to explain how those sketches differed.

She replied, "Wax paper . . . more higher than foil. Glass is spread out more."

"Suppose I want you to explain this to me later. Can you?"

"Yeah."

"How about next year?"

"No."

"What can you do to help me or someone else understand what this shows?" she was asked. However, she did not respond.

"What can you do now?"

"Write down by each picture," she said.

"Would you do that?"

She then proceeded to write down the types of surfaces the different drops were placed on.

"Anything else I need to remember?" she was asked, but again she didn't respond.

<u>Figure 4-2b</u>. Tammy's Record - Additional Perspective Following Leading Probes

"Why is this drop different from this one?" she was asked as the investigator pointed first at one drop then a second.

"Oh, the number of drops," she said and then proceeded to write in this information. When she had finished, the completed record looked like Figure 4-2c.

The first part of the sketch is done in good proportion and fairly represents the actual size of the drops. She has arranged them in her sketch very much like the way they were arranged on the surfaces. Although the second part of her sketch shows more convenient arrangement, she did not include the number of drops per heap, so one must infer with difficulty which is which.

The portrait painted of Tammy's observation competence indicates that it is low and very much like the typical grade six student. While she does appear to possess certain traits which could develop into favourable competencies for making observations, she does not have a clear conception of how to proceed with very simple experiments or activities. In response to leading probes she is able to demonstrate more ability to make and report observations.

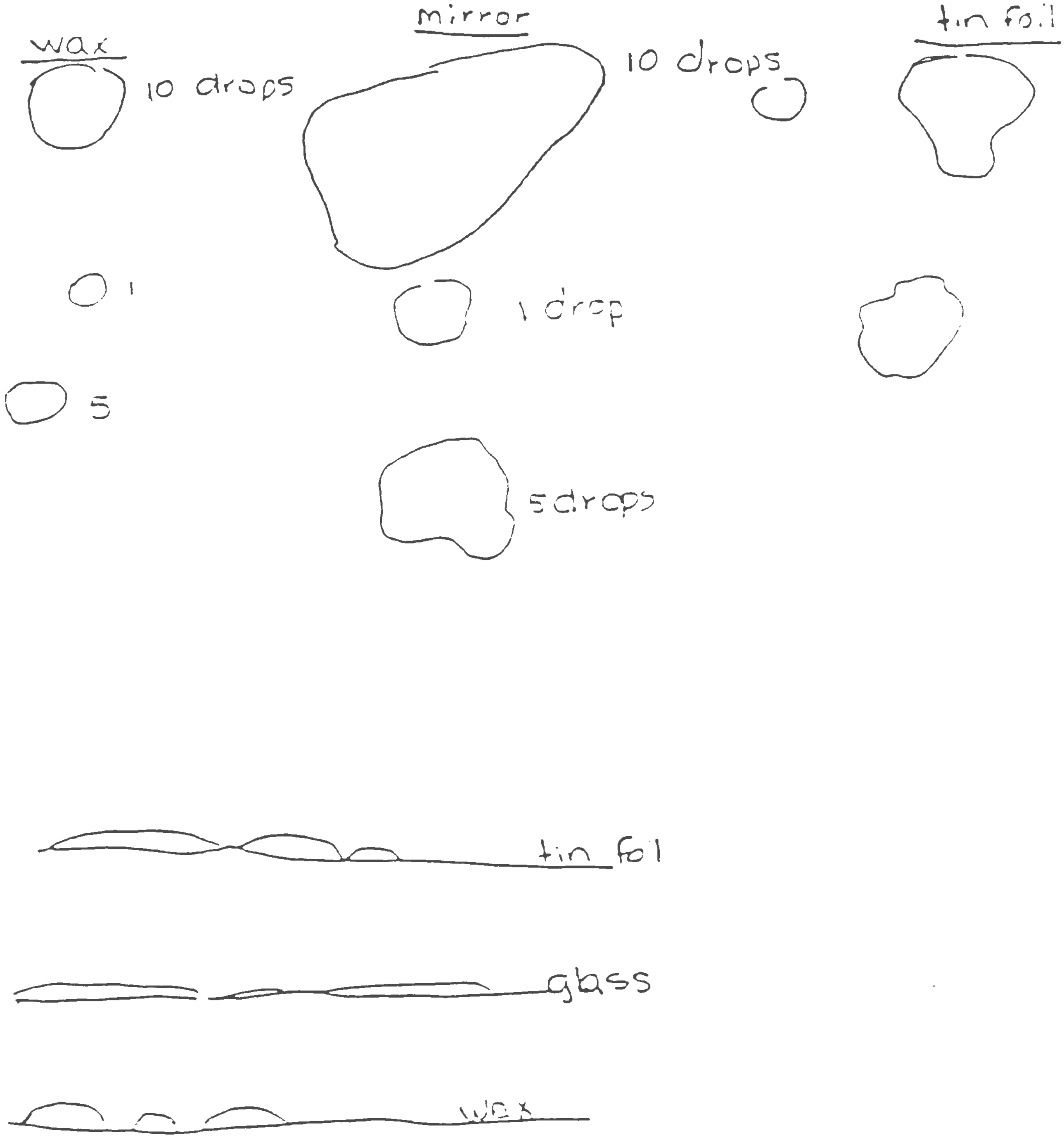


Figure 4-2c. Tammy's Record - Supporting Information Following Leading Probes

<u>Nancy -- An Above Average Observer</u>

Nancy performed very well compared to the group as a whole and indicated excellent observation competence. Throughout the interview her responses, together with her behaviours, indicated that she has extensive general knowledge and she is able to bring this knowledge to bear as she carries out science activities.

<u>Alertness</u>. Nancy is quite alert when making her observations and is quick to notice any unusual events that are relevant to the observation she is making. For example, during the Heads and Tails Capacity Activity she suddenly interrupted her activity.

"Oh!", she exclaimed.

When asked what happened she replied, "Two came out," referring to the fact that two drops had squeezed out together.

"So how did you count it?" she was asked.

"That makes twelve now," she replied, indicating that she was accounting for the unexpected flaw in procedure.

During the Different Surfaces Activity, she was about to place a drop of water on the tin foil when she suddenly moved the dropper to a different position on the foil. She was asked, "Does it matter where you place the drop?"

She replied, "Yeah . . . if there's a bumpy part."

Most students had trouble making the intended observation in the Falling Drops Activity. However, Nancy quickly stated, "It went out and came back like a rubber band."

Even though she lacked the vocabulary to use words like disperse, or contract, or cohesion, she did notice immediately what had happened and used her own vocabulary to describe it.

In the course of the same activity she was required to switch to a longer ruler, a metre stick, because the distance to be measured had gone beyond the range of the 30 centimetre ruler. She selected a metre stick that had a piece broken off the end.

As she picked it up she looked at it and said, "Looks ragged. Starts at fifteen there," but proceeded to use the ruler anyway.

"So you're going to use that one are you?" she was asked.

"There at forty-seven," she replied, not really responding to the probe, and not accounting for the missing centimetres.

This was most uncharacteristic of her more typical behaviour, because she was for the most part quite alert as she made her observations.

Technique. Nancy appears to be skilled at observing the sorts of things observed and is able to understand or decide on an observation technique to be used. In the Heads and Tails Capacity Activity, even though she didn't grasp the technique immediately, the portion of the interview which follows illustrates that she can handle the procedure with minimal guidance.

She was asked, "Which will hold more water? Head or tail side?"

"So, put one on each? Keep going until . . .", she answered leaving her statement unfinished as she continued adding drops to each dime.

"It looks like the tail side."

"How do you know if you can't get any more on?" she was asked.

"I'll try," she replied, and continued adding drops.

"Now, which has more?"

"Can't really tell," she answered. "I'll see if this one can hold more."

"So, how will we know which holds more?" she was asked.

She answered, "Keep going until one overflows."

In this activity it became necessary for her to start over because the water spilled inadvertently. Before starting over, she picked up the two dimes and dried them with a paper towel. She was asked why she had done this. Her response was that, "It might hold a little bit more," indicating she understood that if the technique is to be fair, then variables such as water left on the dimes can interfere with the results.

In the Magnifying with Drops Activity, she was asked how she could tell if the size of the image had changed. "What do you look at?"

Her response, though not well stated, shows that she had a method. "Sort of like the length and all the other letters," she responded.

This response indicates that she is aware of how the dimensions of the letter being observed are seen in relation to the surrounding print as a basis of comparison.

Such competence was evident throughout the series of activities and there seemed to be no cases where she experienced difficulty with observing the sorts of things to be observed or with the technique to be used.

Theoretical understanding. Nancy has good theoretical understanding about the world around her and she is able to draw on this understanding. This was demonstrated in various instances.

In the Heads and Tails Capacity Activity she would sometimes have the dropper held a few centimetres above the dime but would then stop and put the dropper much closer to the dime before letting the drop out. She was asked why she did this.

She replied, "It might overflow. If it hits hard."

Though a seemingly very obvious piece of information, there were a number of subjects who didn't seem to grasp this. They had their testing procedure disrupted by the impact of the falling drop, and didn't realize that nothing had been determined as far as the testing procedure was concerned.

In the Different Surfaces Activity, she was drawing sketches intended to show how the drops differed. She was asked, "How is this one different from that one?" as the investigator pointed first at one sketch and then another.

She misunderstood the question and her response indicates that she was trying to explain why they differed rather than how they differed. She said, "Maybe because of the surface. Seems like that one sticks and the other one slides so that one builds up."

In the Falling Drops Activity, she was having trouble getting the drops to spread and contract and realized that she would need to go quite higher after she had made several trials. "I don't get it," she said after continually higher attempts. "I have to go on up there to make it spread."

She then stood up to reach to a higher level because she understood that if it were to work she would need additional height.

In the Tiniest Drops Activity, she was attempting to make drops that were of different sizes. This is possible if the dropper is tipped on its side or is held close enough to the surface to allow the drop to adhere to the surface before it comes completely out of the dropper. When asked how you can make a smaller drop she replied, "If you could change the tip." This seems to indicate that she has a good conception of why all the drops from one dropper are the same size and realizes that different size tips will create different size drops.

Such theoretical understanding seemed to be constantly present throughout the activities and seemed to contribute to her performing well with other competencies. Because she had good theoretical understanding she knew when to be alert. Because she had good understanding, she could more easily grasp a technique and understand how it applied. Indeed, theoretical understanding would seem to be a major contributor to her observation skill generally.

Preconceived notions. There was very little data to indicate whether or not Nancy is influenced by preconceived notions. On the one hand, she displayed competence here, but in another situation she was less competent. For example, in the Magnifying with Drops Activity, the question typically posed by the investigator was, "So, what do you think will happen when another drop is added?" The tone of the question, along with the noted effect of adding one drop, consistently led students to expect the image to be larger. So it was with Nancy who replied, "It might look bigger."

After adding another drop though she stated, "Not much. It looks like it's getting smaller again."

After adding another drop she then maintained that, "It looks like it's staying the same size."

It would appear that she is prepared to go against her expectations when the observation warrants it.

In another case however, the opposite was noted. She seemed to be unduly influenced by a preconceived notion in the Falling Drops Activity. As she was testing higher and higher levels for wax paper, she stated, "I'll have to go up there to make it spread." She then stood up so she could drop the water from a higher level. "It won't spread," she stated. Then, after noting that the drop was not spreading on wax paper she seemed to report out of desperation, "There. It's 84 centimetres."

There were no other instances that could be construed as being relevant to Nancy's being influenced by preconceived notions, so there is no strong indication one way or the other to indicate how she fares generally with this competency.

Precision of technique. Nancy displays good competence in choosing and using an appropriately precise technique. She knew that a technique would be required to keep track of the water on the head and tail sides of the dimes when trying to determine which side would hold more.

For example, when the activity commenced and she was asked which side would hold more, she replied, "So, put one on each. Keep going until . . ." Her unspoken words seemed to imply that she had a conception of how to continue.

As the activity progressed she was later asked, "Can you tell how much water you put on each dime?"

"Count the drops," she responded.

She had also mentioned the possibility of measuring how much water was gone out of the dropper but since there were no markings on the dropper she was led to opt for counting drops as the more viable option. It would seem that Nancy not only understands the need to use a precise technique but she is also quite imaginative about alternative means of making similar or related observations.

Access. Nancy is quite adept at securing good access to what she is observing. She positions herself well by leaning over, squinting with one eye, or moving things about so that she will see it better. At one point in the Closest Drops Activity, she commented that, "You gotta look right down over it," indicating that she understands that a better view is offered that way than from a side view where it would be difficult to note the space between the drops.

In the Different Surfaces Activity, while looking at the drops on glass, she noted that, "Can't really see on that 'cause it looks like it's double." She was referring to the image of the drop in the glass which interfered with her being able to see clearly the size, shape, and outline of the drop.

It required considerable leaning and squinting before she finally finished sketching and proclaimed, "There, that's the best I can do with that one," indicating that although she had reported her observation, she understood its limitation due to poor access.

Nancy takes ample time for making her observations and will make repeated observations to be sure of what she has observed. For example, when measuring the height at which the spreading and contracting effect is noted when a water drop falls on tin foil, she made many repeated observations at the same height to make sure she had observed properly or had not missed something. This was typical of her behaviour throughout the activities and is indicative of her understanding of the need to verify what is observed.

<u>Instrumentation</u>. Nancy has good facility with instrumentation, which in this case refers to rulers for the most part, and understands very well how it is used. She very adeptly made a measurement with a 30 centimetre ruler to report from what height water drops were falling, and switched to a metre stick when the distance to be measured had extended beyond the range of the smaller ruler. When she first picked up the metre stick she glanced quickly at it and turned it over. When asked why, she said, "I had it that way," and pointed to the 100 centimetre mark on the ruler. It didn't take further probing to infer that she knew just which end of the metre stick should be used.

However, in the incident that was referred to earlier when discussing her alertness, she had noticed that the ruler had 15 centimetres broken off it, had commented on it, but then had proceeded to use the ruler without accounting for the missing portion. This behaviour was uncharacteristic of her overall proficiency with making measurements but it does serve to illustrate that even the most competent student observers have something to learn.

Precision of report. In reporting her observations Nancy tends to report them accurately based on the method used, but no more precisely than the method warrants. For example, when reporting heights in the Falling Drops Activity, she reported that it "seems to happen at about 13 or 14 centimetres." This was as precise as could be expected for the activity and the qualifier "about" indicates that she is aware of the limitations of both the observation and the measurement.

A characteristic she displayed on a couple of occasions was to report precise measurements when no instrument was used. She was comparing the distance between drops of water on tin foil and on glass and observed that, "They're three or four . . . no, two millimetres apart." This precision, though not required and being only estimates based on her judgement of a millimetre, illustrates the confidence she has in her knowledge of measurement units.

Well-made record. Nancy's competence in reporting her observations in a well made record is best displayed when there is the benefit of leading questions to stimulate her. When making a record in the Different Surfaces Activity, she didn't write down any information to indicate which drawings were for which surface, nor did she indicate how many drops of water formed each heap. She had been instructed that the records were for me, the investigator, and that they would need to be clearly done and in detail because I would have to study them later and would need to understand what they were showing. However, she indicated that she had completed her record with Figure 4-3a.

When asked, "Will you be able to explain to me weeks later just what those drawings show?" she replied, "Probably not."

"Would you explain to me now what differences your sketch shows?" she was asked.

"The distance they spread out and how they were built up," she replied.

"How does your sketch show that they are built up?" She was unable to reply to this question and she was then asked, "Is there a way to look at the drops to see how they are built up?"

"Sideways," was her reply.

Figure 4-3a. Nancy's Record - No Probing

She then proceeded to go through the motions of viewing the drops from the side and drawing them from that perspective. By now her record looked like Figure 4-3b.

She was asked, "Will you be able to explain to me in a few weeks what the drawings show? Will you remember which is which?"

She replied, "I don't think so."

"Do you think I'll remember what all this is showing when I look at it much later?"

"Probably not," she replied.

"What can you do now to help us remember?" she was asked.

In response to this probe she replied, "One way is to put down W under it and T under the tin foil pictures."

She proceeded to mark the letters under the appropriate locations as shown in Figure 4-3c.

"I'll also need to know why the drops are different sizes."

"Oh, right. I'll put numbers over it," she said.

Although the record was reasonably neat and organized and in good proportion to the actual drops, it had taken some suggestive probes to get her to produce them. The necessity of completing detailed records with supporting clarification which would help another party interpret the records had not been immediately apparent to her. However, the ease with which she was led to produce more detailed

Figure 4-3b. Nancy's Record - Additional Perspective Following Leading Probes

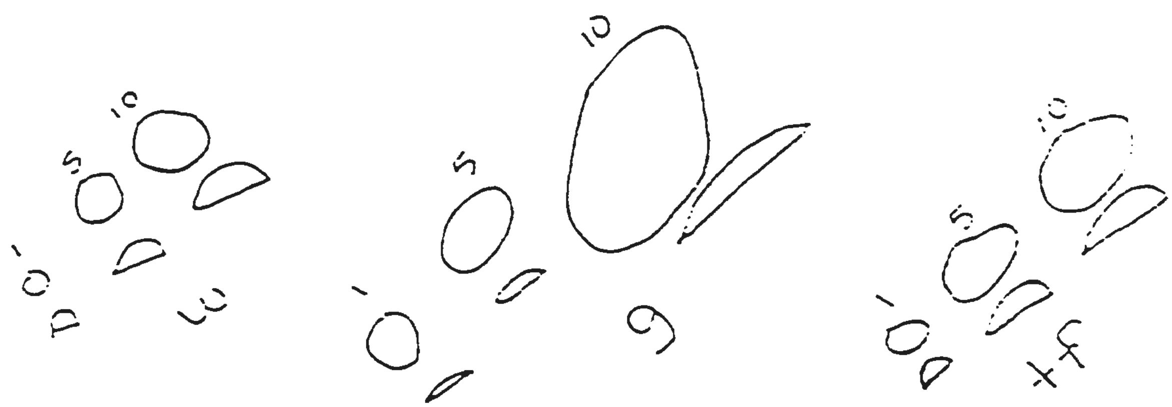

Figure 4-3c. Nancy's Record - Supporting Information Following Leading Probes

records indicates that she is at a stage where building on this competency should be easy for her.

The description of Nancy presents the impression that she has a high degree of skill for making observations but is somewhat less skilled in making her reports of observations when those reports require a record to be made. While she does falter from time to time, she performs with a much greater competence than the typical grade six student. She possesses a broad knowledge base which she brings to bear at all times as she carries out activities and results in her being a skilled, alert, competent observer.

<u>Stephen -- A Below Average Observer</u>

Stephen is a very poor observer. Throughout the course of the interview he continually gave unfavourable responses to questions or else did not formulate any response. It was mainly in response to leading questions that he was able to demonstrate some degree of competence, but for the most part he performed at a level well below that of the typical grade six student. It was very difficult to get Stephen to talk and many of his responses were simply yes or no utterances. When he did try to elaborate he talked very lowly and unclearly. He was very clumsy and awkward with the materials and was very uncoordinated in manipulating the various pieces of equipment.

<u>Alertness</u>. Stephen is not at all alert to the situation at hand. As he conducted the test in the Heads

and Tails Capacity Activity, he bumped the table and caused the water to spill off the head side. He was asked what had happened.

"Head can't hold any more," he replied and started to test the tail side.

No attention was given to the fact that the bump against the table had caused the water to spill and he had to be led to testing the head side again.

In the Different Surfaces Activity, he didn't express any concern that the drops had all run together to form a puddle due to bumps on the tin foil. Instead, he was going to sketch them that way until he was told that he must sketch the drops separately. Such lack of alertness (or thinking or caring) was evident throughout the entire interview and made the complete procedure very difficult to manage.

Technique. Stephen displayed very little skill at observing the types of things observed in the activities and showed almost no skill with the various techniques used. He needed to be told to count the drops that were being placed on each dime in the effort to determine which side could hold more, and was hampered considerably by the fact that he didn't use the dropper properly. He would squeeze too hard and get several drops or a stream of water.

A portion of the interview went as follows after he had started over following the table bumping incident.

"Which holds more water? Heads or tails?"

"Tail," he replied.

"How do you know?"

"'Cause tail got more water than head side," he answered.

"And the head can't hold any more?" he was asked, because the head side had not been spilled.

"No," he said.

"How do you know?"

"It'll run over," he answered.

He was told to place more water on the head side and was again asked which had more water.

"Tail," he again replied.

"How do you know?"

"It's the same thing. Put little squirt on each."

"Can you put more on?"

He then added some more water.

"Now, which has more?"

"Tail," he maintained.

"How do you know?"

"It's higher," he answered.

This type of procedure continued until he was finally told that a good way to keep track of the water was to count the drops being placed on each dime. He did this and finally the head side broke and spilled over.

"Do you know which can hold more?"

"Head," he responded.

"But the head is spilled isn't it?"

"Yes."

"So which is going to hold more water?"

"Tail," he finally declared.

This type of discourse was quite common throughout the activities with Stephen and it often took extremely leading questions to get either a correct or favourable response from him. In many cases he had to be specifically guided, directed, or told in order to keep the activity from completely stalling.

Theoretical understanding. His theoretical understanding of the things he observes is extremely limited. Within the activities he said or did very little to indicate competence. He showed no realization that bumping the table had made the water spill off the dimes, he didn't seem to realize that the impact of the falling drops could cause the water to spill, and he hadn't indicated, except when probed with a leading question, that ridges or rough places on tin foil could cause the drops to run or change shape. When this had happened the investigator had asked if it made any difference where on the tin foil the drops were placed.

"No," he had replied.

The investigator then pointed at a couple of places on the tin foil and asked, "How about if you put the drops there or there?"

"The drops'll run down there," he said, indicating that he does know what will happen although not drawing on that information freely.

At another time in the Heads and Tails Capacity Activity when he was starting over at one point the investigator asked, "Is there some water on the coins already?"

"Yeah," he answered but didn't bother to dry it off.

"Does it matter?" the investigator then asked.

"Doesn't matter," he responded.

No matter what the activity, Stephen demonstrated little understanding. This lack of basic knowledge, or failure to apply basic knowledge, greatly interfered with his ability to make any observations well.

Preconceived notions. Although little data was available regarding the effect of preconceived notions, the tendency noted was that he reports observing what he expects to observe. However, a clear understanding of what Stephen was thinking is clouded by the responses that he often gave. For example, the following interchange comes from the Magnifying with Drops Activity.

"Did anything happen to the number?"

"It looks a bit bigger," he answered.

"What do you think will happen when another drop is added?" he was asked.

"Get bigger," he replied.

When the drop was added he was again asked, "What happened?"

"Got bigger," he said.

"What if we add another drop?"

"Get bigger," he said.

Another drop was added.

"What happened?" he was asked again.

"Got bigger," he said.

This happened a couple more times and the investigator sensed there might be some misunderstanding.

"What's getting bigger?" Stephen was asked.

"The water," he replied.

"Is the number getting bigger?" the investigator asked.

"Yeah," Stephen replied.

Stephen was then asked to add another drop of water.

"What happened?"

"The water got bigger and the number too," he replied.

Finally, the investigator had Stephen place one more drop and asked, "Is the number still getting bigger?"

"No," Stephen answered.

"What's happened now?"

"The water's gone all over it."

"What about the number?" the investigator asked.

"Smaller," was his reply.

"How small is it?"

"Small as the other -- there," he answered pointing at other text.

It almost seemed as if Stephen were intimidated by the whole situation and, in his desire to be right, had been reporting what he suspected might be the response the investigator wanted to hear. Even if this were the case, however, he was still denying the obvious as he continually gave unwarranted reports and responses.

In the Falling Drops Activity he was to find out whether the spreading and contracting effect of a water drop falling on wax paper is like the effect on tin foil, and, if so, at what height the dropper must be held to make it happen. He seemed to understand that it was not happening from low heights as it had on the tin foil and continued to try higher and higher levels.

"Did it happen?" he was asked.

"No."

"Then?"

"No."

"How about that time?"

"No."

He kept going higher and the next time he was asked if it happened he replied "Yeah."

"Find out how high."

He then measured and reported that it was at 48 centimetres. The fact remains however that this effect just doesn't happen on wax paper. It would appear that since Stephen had been asked to find out at what height the effect can be noted, he felt compelled to report something and was prepared to give a false report because he felt that was expected.

Precision of technique. Stephen had to be told in practically all cases which technique to use, so it cannot be said that he chooses as precise a technique as is appropriate. What can be said, however, is that when using a technique, he does not use it precisely. When counting drops that were placed on the dimes he couldn't keep track of the number of drops being placed there because of the spurts of water that would come out, but still he kept on counting, seemingly oblivious to the flaws in handling the technique properly. When attempting to find out at what height the spreading and contracting effect of a drop falling on tin foil is first observable, he was very confused about the technique to use. He tried the dropper at a very low level and correctly reported that it didn't happen.

"So what will you do next?" he was asked.

"Go higher," he correctly replied.

"Why should you try higher?" he was asked.

"To test it again," he said.

"But why do you need to go higher? Why not lower?"

"Squirt water out," he replied, without explaining why it was necessary to test from a higher level.

After a couple of further trials he was asked, "Has it happened yet?"

"Yes," he answered.

"At what height?"

"About that high," he said placing the dropper at a point in mid air.

"We need to know at what height it first happens," he was reminded.

He then took the ruler, stuck the dropper at a random point and measured how high it was.

"Fourteen centimetres," he reported.

"Is that where it first happens?" he was asked. "What's the lowest?"

"Down there," he answered and placed the dropper close to the surface and measured again.

"Two centimetres," he said. "Just one."

Stephen seems to place little importance on precision and didn't give any response that would indicate he has appreciation for the need to be precise.

Access. In gaining access to the thing observed Stephen again demonstrated a low degree of competence He made no effort to position himself advantageously for best viewing and in many cases made cursory inspections of what was to be observed. For example, when sketching drops of water on tin foil, wax paper, and glass he would lean down to take a quick look then sit upright to draw from memory what he had observed. This makes it quite difficult to get the proper scale to his drawings.

The procedure of getting him to view the drops from a side view was demanding. After he had sketched the drops from an overhead view, he was asked, "Is there another way to look at them?"

"Sideways," he answered. He tilted his head slightly and looked at the drops.

"Can you see them that way?"

"No," he answered.

"How about if you get down lower?" he was asked.

He tilted his head again but still didn't get lower for a better view.

"Can you draw them from the side view?" he was asked.

He didn't respond, nor did he do anything, so the investigator showed him how to peer at the drops from the side and how to sketch them. He was then directed to continue with the drawing.

He began drawing, however he still had not positioned himself low enough to get a proper side view so he was prompted with another question.

"Are you sure you have a good view from there?"

"Yeah," he answered.

The investigator then glanced at his out of proportion sketches and asked, "Are they piled up that high?"

"No," he answered.

"Would it help if you got closer?" he was asked in the hope that a better view might result in more accurate sketching.

"Yeah," he replied.

"How are the pictures different from the side view?" he was asked when he appeared to be finished.

"Height. They're higher," he replied.

Stephen had to be continually encouraged to make repeated observations. As can be seen from the foregoing dialogue, his responses to questions were mainly yes or no. This is probably because of the nature of the questions themselves for with Stephen most questions were guiding type questions to try to get him to perform the activity properly. It was very difficult to get any elaboration from him on anything. He did not express anything that indicated he has some conception of what good access is and the advantage it would afford for better observing. And he certainly did not volunteer to secure better access.

<u>Instrumentation</u>. It cannot be said that Stephen displays strong ability with instrumentation although this was the only condition where he came close to the sample average. When required to measure the height that drops were falling from the dropper he was able to choose a

suitable ruler with the proper range. However, he was awkward and uncoordinated in holding the dropper properly with one hand, trying to position the ruler with the other and then reading the required information from it. As indicated from the simple discourse which follows, he does understand the purpose of some rulers being longer than others.

After trying at repeatedly higher levels to get the spreading and contracting effect when a water drop falls on wax paper, he had gone beyond the range of the thirty centimetre ruler. He laid the ruler down and picked up the metre stick.

"Why are you taking that ruler?" he was asked.

"Longer," was his response.

Basically he had been able to demonstrate that he does have an understanding of how the instruments, rulers, work but he is extremely unskilled in using them.

Precision of report. Stephen did such a poor job of using a technique or making measurements that there was very little to report and it was obvious, though apparently not to Stephen, that his reports were imprecise and inadequate and could not be taken seriously. Still, he reported them as though they were accurate.

As indicated previously, Stephen would do such things as place the dropper randomly at some point in mid-air and make a measurement without rechecking to see whether the dropper was in the proper position. Also, he reported that the tail side of a dime could hold more water even while he was still adding more water to each dime. Behaviours such as these were quite typical of the way he handled all the activities.

Well-made record. Stephen shows very poor ability in reporting observations in a record. His sketches of drops of water on tin foil, wax paper, and glass were out of proportion, and he included no clarification of what the pictures showed. However, in response to leading questions he did show that he understands such information can be useful.

After he had drawn his sketches of the water drops, he was asked, "Could you tell me later what your drawings show?"

"Yeah," he answered.

"How about weeks or months later?"

"No."

"What can you do to help you remember?" he was asked.

"Write down," he said.

"What would you write down?"

"Wax paper," he replied.

"What else?"

"Ten drops, like that," he said as he wrote down the required information. However, he didn't complete the required information for the other drops and his record looked like Figure 4-4. As can be seen, there is little information provided other than for the drops on wax paper.

His initial attempt had been to produce a sketch of the drops from above and he had to be led and directed to produce side views. This was not untypical, for most students did not think to draw a side view of the water drops. However, in Stephen's case, once the sketches were done and he indicated the necessity of writing in information to clarify his drawings, he still did not put in much of the required information. As a result, his record remains unclear and is meaningful only to someone closely familiar with the activity.

Stephen demonstrates very poor observation ability. He is hampered by an extensive lack of theoretical knowledge, is very unalert to the course of the activity, and is hindered by his own awkwardness in handling materials. A great deal of improvement is needed before he can be as proficient as the typical grade six student.

## Summary

The data presented in this chapter has been partially quantitative but the bulk of it has been qualitative. The quantitative data provides summary type information and relates general trends in performance of grade six students as observers.


Figure 4-4. Stephen's Complete Record Following Leading Probes

Among the findings are:

1. Students generally are not good observers. They indicate competency in their observation ability in less than half of the occasions to do so.
2. When probed with leading questions students display a higher level of observation competence.
3. Theoretical understanding of what is being observed is an area of weakness. Although students appear to be as competent in this area as with the other competencies, there is no leading probe effect.
4. The weakest area seems to be in reporting observations in a well-made record. Students displayed very poor ability to present the results of their observations and didn't understand how to arrange their information along with the clarifying details that would be necessary to help another party understand the record. There was a strong leading probe effect with this competency.
5. There is no appreciable difference in the observation ability of the sexes.

The qualitative data took the form of describing three specific students and the typical student. The description of the typical grade six student was developed from the averages of students' positive response rate. Those averages provide an indication of how well the student fares with the various competencies. Typical student behaviours were offered to illustrate how the student's level of competence manifests itself in the student's performance.

The three individuals that were described were chosen because they represent three levels of ability. They are an average observer, an above-average observer, and a below-average observer. The descriptions of each indicate specifically in what areas the students have strengths and weaknesses. This was supported with actual quotes from student interviews along with examples of their physical behaviours. It was noted that the above-average observer demonstrated competence primarily in response to non-leading questions thus making it unnecessary to pose many leading questions. The average observer demonstrated some competence in response to non-leading questions, however, it frequently took the stimulation of leading probes to get her to demonstrate competence. The below-average observer was unable to demonstrate much observation ability in response to either non-leading or leading questions and required continual guidance with all aspects of the observing process.

# CHAPTER 5: SUMMARY AND IMPLICATIONS

## Purpose of the Study

The purpose of this study was twofold: (a) to develop a technique suitable for assessing observation ability of elementary school children; and (b) to assess the observation ability of a sample of grade six students. Tests which are used to provide a measure of observation ability are available, however, they do not specify in just what areas the student has skill or lacks skill. The purpose of the testing technique in this study was to compensate for this weakness and to provide descriptions of student observation ability so that specific areas of strength or weakness could be detected.

## Motivation for the Study

For a number of years a popular aim of school science programs has been to promote acquisition of skill with the science processes. Observation has long been recognized as a very important science process and is considered fundamental to the progress of science. Poor observation ability results in poor science.

Observation ability has also been recognized as one of the bases for critical thinking. In the activity of thinking critically one of the main sources of information is what is gained directly through one's own observations or the reports of observations that are gained from others. If one is to be a critical thinker then good observation skill is paramount.

Skill at observing is believed by many to be a transferable skill. The competencies that one uses to make observations in school science are essentially the same competencies that are used in thinking critically. Similarly, those same competencies are used when witnessing an accident or when trying to decide for whom to vote. Observation skill is believed worthy of being taught not only because of its relevance to science but also because of its relevance to everyday affairs.

Since observation ability is recognized as such a vital skill, worthy of being taught, it follows that we should wish to have effective means of evaluating or assessing it. If teachers are to try to make students into more competent observers, then they need to be able to evaluate the results of such instruction and determine specifically in what areas they have been successful and where they have not. Similarly, programs or instructional strategies can be evaluated as to their effectiveness in promoting this skill.

## Method

The study required students to work through a series of science activities in which they were to make observations and report them to the investigator. They were instructed that as they performed the activities they were to say anything they wished about what they were thinking. They were also informed that they would be asked simple questions about what they did, what they said, or what they were thinking. Two main categories of questions or probes were used by the interviewer: non-leading and leading types of questions. Use of this method to conduct qualitative research has been elaborated in the literature by Ericsson and Simon (1984) and it has been applied in many research situations such as by Finley (1986), Lavoie and Good (1988), and Norris and King (1984).

A model of observation competence proposed by Norris (1984) lists a set of conditions that define what a good observer is like. Norris lists conditions such as the observer being alert, having theoretical understanding, or using a precise technique as observations are made. He also lists conditions such as the observer reporting the observations in a well made record. This model of observation competence provided the criteria for judging students' observation ability in this study.

The responses that the students gave to the probing questions were rated on whether or not they provided evidence related to observation ability. The ratio of positive responses to all responses made relevant to a particular competence was taken as the index of success that the student has with that particular competency. Average scores for each competency and an overall index of performance for the average grade six student were computed. Those figures were then used to judge what level of skill the typical grade six student possesses and examples of how this skill is manifested were selected from the interviews. The result is a description of the typical grade six student's observation competence.

Three students were selected -- an average observer, an above-average observer, and a below-average observer -- to be described in detail according to the level of skill they displayed with the various competencies. Specific behaviours and verbalizations from the interviews were selected to exemplify how each performed.

## Findings

It was found that the data acquired on individual students was very useful in generating the qualitative descriptions that were sought. It was possible to evaluate the observation ability of specific students and point out where their strengths and weaknesses lie.

The accumulation of the data was useful in determining what level of competence is possessed by the typical grade six student and in what areas there are general strengths or weaknesses. It was found that, overall, grade six students are not very good observers when left to their own devices. However, if they are stimulated with leading probes they perform at a much more satisfactory level. A major problem faced by students is a lack of theoretical knowledge about what they are observing. This interfered considerably with their ability to observe well and resulted in their making incorrect observations or no observations at all. Leading questions produced no noticeable effect with this competency.

The greatest level of competency was shown with the instrumentation condition. However, it was speculated that this possibly reflects competence with rulers more so than with instrumentation generally.

The weakest area for students when non-leading questions were asked was in reporting their observations in a well-made record. Students consistently created haphazard, incomplete records without the clarifying details that would be needed for later understandi and interpretation by the observer or by a different person. There was, however, a strong leading probe effect and students, with suggestive stimulation, can produce much improved records.

A last finding was that there is no appreciable difference in the observation ability of the sexes.

## Implications

### Assessment of Technique

The technique used in this study to assess the observation ability of grade six students was experimental. Although its basis is found in the work of other accepted methodologies, there is no report of its having been duplicated elsewhere. One of the unstated aims of the study was to assess the effectiveness of the method. Some strengths and weaknesses became evident as the study was carried out and these need to be commented on.

The method itself is very time consuming and can be used with only one student at a time. Additional time is required with interpretation of the interviews for it is unfeasible to conduct the interviews and do the scoring at the same time. The technique is extremely burdensome if one wishes to assess a large number of students.

It is not possible to study all the aspects of the Norris model of observation competence, so the profiles of individuals' observation competence are not complete. Some characteristics of good observers could not be assessed with the types of activities used in this study. That is not to say that there are no activities that could be suitable for assessing those other competencies, however it is unlikely that factors such as conflict of interest or emotional state are typically inherent in students' classroom science activities in the sense in which Norris conceives of them. Some other approach would be required to assess those competencies.

All students are different and have their own habits, personalities, and idiosyncrasies. These basic differences make some students less suitable for this type of assessment than others. Some students don't talk much due to shyness or perhaps feeling threatened or intimidated, while others do not express their thoughts clearly. This often made it difficult to make inferences about the observation ability of the student and might possibly have led to the wrong interpretation. As a result much of the conversation of the subjects was discarded for its lack of clarity, poor articulation, or lack of relevance. This problem would probably be lessened with older subjects, but with grade six students it does detract from the fullness and richness of the data.

The limited number of activities that could be used is of course a difficulty. All such studies face this problem for only a sample from a seemingly infinite pool of items can be employed. The activities for this study all related to the use of water drops. It might have been possible to get a better sampling of student competencies if a more diverse set of activities had been used. Notwithstanding the earlier justification for deciding to use the activities that were chosen, there would seem to be more merit in using a more varied set of activities. Students could have called upon different theoretical understandings, more varied techniques, more varied instrumentation, and the like. The better the sampling, the more confidence we can place in the results.

Despite those weaknesses, there are some positive things about the method used. It does give the opportunity for the investigator to gain some understanding of the types of information the student is dealing with as observations are being made. The method does allow for broad description of individual students and also allows us to understand what we can expect of the typical grade six student. Rather than providing just a number, which tells how successful the observer was, it additionally allows us to see what specific competencies the student possesses. It doesn't merely tell us how well the observer performed, it also helps us to understand why.

Refinements to Norris Model

The Norris Model of observation competence was developed from earlier work by Ennis in critical thinking. It was not developed from empirical research with students to ascertain what factors indeed do contribute to observation competence. The outcomes of this study suggest that the model is incomplete and there are additional conditions that contribute to good observing.

The first such condition that contributes to observing well is that there be some purpose in the observing process. Students need to know specifically what they are to observe. For example, a common problem occurred in this study when students were told to observe the effect of letting a drop of water fall onto tin foil, and, subsequently, to find from what height this effect can be noticed. Students were unsuccessful in observing that the water drop will spread out upon impact and then contract inward into a smaller drop. Most students reported that they didn't observe anything while several observed that it made a clicking sound when it hit the foil. But when told to watch the shape of the drop and observe how it changes they were much more successful.

It may be useful to think about the purpose of observation in an everyday type of situation. Suppose a person were witness to an automobile accident. That person may or may not be a good witness if called into court. But if the person had realized at the scene of the accident that a court appearance was a possibility, then more purposeful observations might have been made. For example, the time, number of cars involved, direction from which cars had come, and the estimated speed of the cars are the sorts of things that are relevant when investigations of traffic accidents are carried out. If a witness thinks to observe with a purpose, then it is likely that more useful and more accurate observations would be made.

It is suggested that an addition to the Norris model would be: In order to observe well an observer should observe with a purpose.

Another condition that contributes to observing well is having good access to the thing observed. This is suggested in the Norris model and some breakdown of examples of what constitutes good access is suggested. Missing from the Norris model is a statement about the importance of the observer having the best possible vantage point for making observations. Most things can be observed from various perspectives and from various distances resulting in varying observations. This factor was made obvious in the activity which required students to draw sketches of water drops. Many of them did not think to position themselves directly above the drops so they could look down upon them nor did they think to position themselves at eye level with the drops so they could view them accurately from a side perspective. The tendency was to view the drops obliquely, unless they were prompted to seek a different perspective.

An observer should be positioned at the most advantageous place for making the required observations and the accuracy of observations is limited to the extent that this condition has been met. Although this condition may be assumed implicitly in the Norris model, it needs to be made more explicit.

A second addition to the Norris model would be made in category 9 which states the conditions of good access: Access is good to the extent that there is a satisfactory vantage point from which to observe.

A third condition that contributes to observing well is that the observer is imaginative about the technique to be used, and which observations to make. For example, in the Heads and Tails Capacity Activity, students were not told immediately how they should proceed. Instead they were left to their own devices. Many students were unsuccessful in deciding on a technique while others mentioned a couple of possibilities, usually involving counting the drops. Three or four students also mentioned the possibility of putting marks on the dropper to indicate how much water had been removed from it. While it may not have been the most workable method, it does indicate the activity of an imaginative mind that considers various possibilities.

Many of the greatest advances in science have been the result of imaginative insight by people making routine observations or carrying out routine investigation. By being imaginative an observer stands a better chance of making more meaningful, more precise, and perhaps more useful observations.

A third addition to the Norris model would state: In order to observe well an observer should attempt to be imaginative about the technique to use and observations to make.

The Norris model acknowledges the importance of making reports in a well-made record. No mention is made, though, of the best form or condition for oral reports of observations. It would be a sensible thing to expect that they should be well presented and in a form that can be understood by another individual or an audience.

Some of the subjects in the study were unable to communicate their observations very clearly and it required some insight, interpretation, and inference on the part of the investigator to understand what was being communicated. This enhances the possibility that the observation reports were misunderstood even though the observations may have been made correctly.

A fourth addition to the Norris model would state that: In order to report observations well an observer should report the observations using clear, concise, and well-articulated language.

It would seem that these four additions to the Norris model would make it more encompassing and complete. Further research might show that there are other additions and refinements to be made.

Educational Considerations

It was found in this study that grade six students are not good observers. They indicated competence in less than half of the opportunities to do so. However, when probed with leading questions, they demonstrated a more satisfactory level of competence.

It is possible that the level of competence displayed by the students is all we can and should expect of grade six students. This is what they are like and no amount of effort is going to alter the fact. If that is the case, then we continue to instruct students as always and at least maintain the status quo.

But the leading probe effect would seem to indicate that students have somewhat more competence than they typically show. When students were stimulated to examine their thoughts or the knowledge they already held to see if it was relevant, they generally were able to display a higher level of observation competence. The task then is to find a way to get students into the habit of closely examining their thoughts and their presently held knowledge to see what is relevant to the task at hand and how it can be applied. Students must learn to do this themselves without being stimulated to do so by leading questions from the teacher. This is a habit we wish them to develop and apply in all their everyday affairs. Science class is merely one place to promote it.

What then does this mean for what is happening in the classroom? There are at least three factors which can have an impact upon making students become better observers. These include: (a) Program, (b) Teacher, and (c) A Model of Observation Competence.

Let's explore them in that order.

Program. The science program followed by the students in this study was STEM Science. It had been revised as Addison-Wesley Science by the time they took part in this study. A great deal of the program is concerned with content and conceptual understandings, but in addition there is a major emphasis on the science processes. Many activities are recommended as exploration and experimentation. If the program is followed, there would be many opportunities for students to practice observing in a school science setting. However, the net effect has been that use of this program has not resulted in students who are competent observers. Assuming that the program has been implemented properly, then the implication is that it is inadequate in meeting one of its important aims as well as an important aim of science education.

Teachers. It is possible that the program is not at fault but that teachers have not implemented it properly. First, we need to find out whether this is so. This will require that a comparative study be done with a school where it is known that the program is being taught properly. It would determine if students from such a setting are indeed better observers. If they are then the solution lies with in-servicing or retraining teachers so that they do implement the program properly. If students from such a setting are not found to be better observers then we are back to the program.

A model of observation competence. It is possible that the basis of this problem, poor observation ability, rests with students. They do not have the critical spirit, the disposition to make use of the competency they possess. If this is so then a major effort must be made to get students to exercise their competency. In short, they need to be made aware of the competencies that a good observer has and encouraged to develop and practice those skills in their school science studies as well as in all other areas. This will require, of course, that teachers first become familiar with the model so that they in turn can instruct students. Students must learn to be alert, to use suitable precision, and to use instrumentation taking into account its working condition. Students must learn to report their observations with the appropriate precision and to make good records when their observations are reported in a record. Once students have come to understand the sorts of things that facilitate good observing, and realize the sorts of things that are relevant to the observing process, then, perhaps, they will be more disposed to exercise the abilities they already have.

## REFERENCES


Addison-Wesley science. (1984). Menlo Park, CA: Addison-Wesley.

Afflerbach, P., & Johnston, P. (1984). On the use of verbal reports in reading research. Journal of Reading Behavior, 16, 307-322.

American Association for the Advancement of Science. (1967-68). Science--A Process Approach, New York: Xerox Corporation.

American Psychological Association (1985). Standards for educational and psychological tests. Washington, DC: APA.

Brandt, R. (1990). On knowledge and cognitive skills: A conversation with David Perkins. Educational Leadership, 48(6), 50-53.

Bredderman, T. (1982). The effects of activity-based science in elementary schools. In M.B. Rowe (Ed.), Education in the 80's: Science. Washington, DC: National Education Association.

Clifford, B.R., & Scott, J. (1978). Individual and situational factors in eyewitness testimony. Journal of Applied Psychology, 63, 352-359.

Crocker, R.K. (1973). Elementary science curriculum study. Toronto: McGraw-Hill Ryerson.

Cronbach, L.J. (1971). Test Validation. In R.L. Thorndike (Ed.). Educational measurement (2nd ed.). Washington, DC: American Council on Education.

Cronbach, L.J., & Meehl, P.E. (1955). Construct validity in psychological tests. Psychological Bulletin, 52, 281-301.

Dale, P.S., Loftus, E.F., & Rathbun, L. (1978). The influence of the form of the question on the eyewitness testimony of preschool children. Journal of Psycholinguistic Research, 7, 269-277.

Dewey, J. (1933). How we think. Boston: D.C. Heath & Co.

Ebel, R.L. (1972). Measurement and the teacher. In Noll, V.H., Scannell, D.P., & Noll, R.P. (Eds.), Introductory readings in educational measurement. Boston: Houghton, Mifflin.

Ennis, R.H. (1962). A concept of critical thinking. Harvard Educational Review, 32, 81-111.

Ennis, R.H. (1980). A conception of rational thinking. In J.R. Coombs (Ed.), Philosophy of education 1979 (pp. 3-30). Normal, IL: Philosophy of Education Society.

Ennis, R.H. (1985). A logical basis for measuring critical thinking skills. Educational Leadership, 43(2), 44-48.

Ennis, R.H. (1989). Critical thinking and subject specificity: Clarification and needed research. Educational Researcher, 18(3), 4-10.

Ennis, R.H. (1990). The extent to which critical thinking is subject-specific: Further clarification. Educational Researcher, 19(4), 13-16.

Ennis, R.H., & Millman, J. (1985). Cornell Critical Thinking Tests, Levels X and Z. Pacific Grove, CA: Midwest.

Erickson, G.L., & Erickson, L.J. (1984). Females and science achievement: Evidence, explanation, and implications. Science Education, 68, 63-89.

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

Ericsson, K.A., & Simon, H.A. (1984). Protocol analysis: Verbal reports as data. Cambridge, MA: MIT Press.

Finley, F. (1983). Science processes. Journal of Research in Science Teaching, 20, 47-54.

Finley, F. (1986). Evaluating instruction: The complementary use of clinical interviews. Journal of Research in Science Teaching, 23, 635-650.

Gagné, R.M. (1963). The learning requirements for enquiry. Journal of Research in Science Teaching, 1, 144-153.

Gagné, R.M. (1965). The psychological basis of Science -- A Process Approach. Washington, DC: American Association for the Advancement of Science.

Glaser, E.M. (1985). Critical thinking: Education for responsible citizenship in a democracy. National Forum, 65(1), 24-27.

Hanson, N.R. (1958). Patterns of discovery. Cambridge: Cambridge University Press.

Hirsch, E.D. (1987). Cultural literacy: What every American needs to know. Boston: Houghton Mifflin.

Holland, J.H., Holyoak, K.J., Nisbett, R.E., & Thagard, P.R. (1986). Induction: Processes of inference, learning, and discovery. Cambridge, MA: MIT Press.

Hungerford, H.R., & Miles, D.T. (1969). A test to measure observation and comparison skills in science. Science Education, 53, 61-66.

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

Lavoi, D.R., & Good, R. (1988). The nature and use of prediction skills in a biological computer simulation. Journal of Research in Science Teaching, 25, 335-360.

Layton, D. (1973). Science for the people. New York: Science History Publications.

Lipton, J.P. (1977). On the psychology of eyewitness testimony. Journal of Applied Psychology, 62, 90-95.

Loftus, E.F. (1979). Eyewitness testimony. Cambridge, MA: Harvard University Press.

Marquis, K.H., Marshall, J., & Oskamp, S. (1972). Testimony validity as a function of question form, atmosphere, and item difficulty. Journal of Applied Social Psychology, 2, 167-186.

Martin, M. (1972). Concepts of science education. Glenview, IL: Scott, Foresman.

McPeck, J. (1981). Critical thinking and education. New York: St. Martin's.

McPeck, J.E. (1990). Critical thinking and subject specificity: A reply to Ennis. Educational Researcher, 19(4), 10-12.

Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. American Psychologist, 30, 955-966.

National Assessment of Educational Progress (NAEP). (1978). Science achievement in the schools: A summary of results from the 1976-77 national assessment of science. Denver, CO: Education Commission of the States.

Nelson, M.A., & Abraham, E.C. (1973). Inquiry skill measure. Journal of Research in Science Teaching, 10, 291-297.

Nisbett, R.E., & Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes. Psychological Review, 84, 231-259.

Norris, S.P. (1984). Defining observational competence. Science Education, 68, 129-142.

Norris, S. (1985). The philosophical basis of observation in science and science education. Journal of Research in Science Teaching, 22, 817-833.

Norris, S.P. (1988). Controlling for background beliefs when developing multiple-choice critical thinking tests. Educational Measurement, 7(3), 5-11.

Norris, S.P. (1990). Effect of eliciting verbal reports of thinking on critical thinking test performance. Journal of Educational Measurement, 27, 41-58.

Norris, S.P., & Ennis, R.H. (1989). Evaluating critical thinking. Pacific Grove, CA: Midwest.

Norris, S.P., & King, R. (1984). The design of a critical thinking test on appraising observations. St. John's, Newfoundland: Memorial University of Newfoundland, Institute for Educational Research and Development.

Paul, R.W. (1985). The critical-thinking movement, a historical perspective. National Forum, 65(1), 2-3, 32.

Perkins, D.N. (1981). The mind's best work. Cambridge, MA: Harvard University Press.

Perkins, D.N., & Salomon, G. (1988). Teaching for transfer. Educational Leadership, 46(1), 22-32.

Perkins, D.N., & Salomon, G. (1989). Are cognitive skills context bound? Educational Researcher, 18(1), 16-25.

Schoenfeld, A.H. (1985). Mathematical problem solving. New York: Academic Press.

Schoenfeld, A.H., & Hermann, D.J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. Journal of Experimental Psychology: Learning, Memory, and Cognition, 8, 484-494.

Science curriculum improvement study. (1968). Berkley, CA: University of California.

Searching for Structure (1981). Toronto: Holt, Rinehart & Winston.

Shapere, D. (1982). The concept of observation in science and philosophy. Philosophy of Science, 49, 485-525.

Siegel, H. (1980). Critical thinking as an educational ideal. Educational Forum, 45(1), 7-23.

Smith, E.R., & Miller, F.D. (1978). Limits on perception of cognitive processes: A reply to Nisbett and Wilson. Psychological Review, 85, 355-362.

Somers, R.L. & Lagdamen, J.M. (1975). The effect of a modern elementary science curriculum on the ability of Filipino children to observe, compare, and classify geometric figures. Journal of Research in Science Teaching, 12, 297-303.

Stanley, J., & Hopkins, K. (1981). Educational and psychological measurement and evaluation (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.

STEM science. (1977). Menlo Park, CA: Addison-Wesley.

Sternberg, R.J. (1983). Criteria for intellectual skills training. Educational Researcher, 12(2), 6-12, 26.

Sternberg, R.J. (1987). Questions and answers about the nature and teaching of thinking skills. In J.B. Baron & R.J. Sternberg (Eds.), Teaching thinking skills. Theory and practice (pp. 251-259). New York: Freeman.

Tenopyr, M. (1977). Content-construct confusion. Personnel Psychology, 30, 47-54.

Thorndike, E.L. (1917). Reading as reasoning: A study of mistakes in paragraph reading. Journal of Educational Research, 8, 323-332.

Thorndike, R.L. (1971). Educational measurement for the seventies. In R.L. Thorndike (Ed.). Educational Measurement. Washington, DC: American Council on Education.