DOCUMENT RESUME

ED 295 833 SE 049 214

| | |
|---|---|
| AUTHOR | Carey, Susan; And Others |
| TITLE | What Junior High School Students Do, Can, and Should Know about the Nature of Science and Scientific Inquiry. Technical Report 88-2. |
| INSTITUTION | Educational Technology Center, Cambridge, MA. |
| SPONS AGENCY | Office of Educational Research and Improvement (ED), Washington, DC. |
| PUB DATE | Mar 88 |
| CONTRACT | 400-83-0041 |
| NOTE | 72p. |
| PUB TYPE | Reports - Research/Technical (143) |
| EDRS PRICE | MF01/PC03 Plus Postage. |
| DESCRIPTORS | Critical Thinking; *Epistemology; Inquiry; Junior High Schools; *Junior High School Students; *Knowledge Level; *Program Descriptions; Science Education; Science Instruction; Scientific Attitudes; *Scientific Methodology; Secondary School Science |

ABSTRACT

One goal of science educators is to help students to understand the nature of scientific knowledge and reasoning. Reported are ideas related to what junior high students do, can and should know about the nature of science and the use of technology in imparting such knowledge. The studies reported in this document are designed to probe junior high students' initial understanding of the nature and purpose of science, the process of scientific inquiry, and to explore whether it is possible to move students beyond their initial conceptions. Study 1 reports the first field trial of a unit designed to replace the typical junior high unit on the scientific method. Study 2 reports the results of a trial in which a teacher used this unit with her seventh grade classes. Appendices include: (1) "Study 1 Written Test"; (2) "Study 1 Clinical Interview"; (3) "Quantitative Analysis of Study 1 Pre/Posttest Clinical Interview"; (4) "Study 2 Written Test"; and (5) "Study 2 Revised Clinical Interview." (CW)

TR88-2





# WHAT JUNIOR HIGH SCHOOL STUDENTS DO, CAN, AND SHOULD KNOW ABOUT THE NATURE OF SCIENCE AND SCIENTIFIC INQUIRY

Technical Report
March 1988

Educational Technology Center
Harvard Graduate School of Education
337 Gutman Library Appian Way Cambridge MA 02138
(617) 495-9373

**What Junior High School Students Do, Can, and Should Know About the Nature of Science and Scientific Inquiry**

Technical Report
March 1988


Prepared by:

Susan Carey
Risa Evans
Maya Honda
Eileen Jay
Chris Unger

Nature of Science Group:

Diane Bemis
Susan Carey
Risa Evans
Mary Greaves
Maya Honda
Eileen Jay
Wayne O'Neil
Sandra Smith
Chris Unger
Al Weinstein



Preparation of this report was supported in part by the Office of Educational Research and Improvement (Contract #OERI 400-83-0041). Opinions expressed herein are not necessarily shared by OERI and do not represent OERI policy.

**Acknowledgements**

The Nature of Science Unit discussed here is the product of the Nature of Science Group at the Educational Technology Center, all members of which are thanked for their many contributions.

The studies reported on here took place during the winter of 1986 and the spring of 1987 at Day Junior High School in Newton, Massachusetts and at Hosmer East School in Watertown, Massachusetts. The Nature of Science Project gratefully acknowledges the cooperation of the Newton Public Schools and the Watertown Public Schools with these studies. We would especially like to thank the following people for facilitating our work in these two schools:

George Buckley, Science Head of the Watertown Public Schools

Joseph Carroll, Principal of Hosmer East School

Mary Greaves, Science Teacher at Day Junior High School

William Jesdale, Principal of Day Junior High School

Robert Kilburn, Science Coordinator of the Newton Public Schools

Sandra Smith, Science Teacher at Hosmer East School

We would also like to thank Diane Bemis, Science Teacher at Watertown High School, for piloting an earlier version of the yeast lessons; and, Al Weinstein, Science Teacher at Cambridge Rindge and Latin High School, for piloting one of the assessment measures.

Finally, we would like to thank the many students at Day Junior High School, Hosmer East School and Cambridge Rindge and Latin High School who participated in this work.

## TABLE OF CONTENTS

## INTRODUCTION

### The Goals of the ETC Nature of Science Group

As science educators we have two goals. First, we hope students will learn the concepts and theories of the biological and physical sciences. Second, we hope they will come to understand the nature of scientific knowledge and reasoning. The Nature of Science Group explores this second goal for the junior high school grades. We have investigated what junior high school students do, can, and should know about the nature of science, and we have explored the possible use of technology in imparting such knowledge.

Typically, the junior high school science curriculum limits discussion of scientific inquiry to the *how* of scientific reasoning and experimentation, neglecting the *why*. At least since the time of the curricular reforms of the 1960's, one emphasis in science education has been to impart the "process skills" involved in constructing scientific knowledge. These skills are extremely varied, and range from such general skills as classification and careful observation -- called "basic process skills" in <u>Science -- A Process Approach</u>'s (SAPA) terminology -- to more complicated and specific skills such as conducting controlled experiments, preparing data tables, and graphing the results of experiments -- called "integrated process skills" in SAPA's terminology. However, the motivation or justification for using these process skills in theory construction is not explicitly treated in the curriculum. We have found this to be particularly true of the presentation of "the scientific method" in current junior high school textbooks, standardized tests, and curricular materials (Carey et al., 1986).

We certainly do not deny the importance of teaching process skills. Junior high school students do not spontaneously measure and control variables, or systematically record data when attempting experimental work. In the current curriculum, students are taught to explicitly state hypotheses, to design experimental tests of hypotheses, to measure, to collect reliable data, and to draw conclusions from these data. Our critique of this focus on the scientific method is not that its points are unimportant, but that its points are incomplete, and thus reinforce misconceptions about the nature of science.

We believe that students must learn to reason critically about scientific knowledge. It is crucial that students understand that the body of scientific knowledge -- the main focus of much of current junior high school science teaching -- is constructed and changing, rather than "the truth." Developing a constructivist view of scientific inquiry and knowledge is generally considered important to the training of future scientists, as well as to a lay understanding of scientific information (National Science Board Commission on Precollege Education in Mathematics, Science, and Technology, 1983). Yet relatively little effort has gone into developing materials that will impart an understanding of the nature of scientific knowledge and the reasoning that supports this body of knowledge -- of the tentative groping toward a deeper understanding of the world, of the cumulative and intellectual process of theory construction. As we will demonstrate, students' initial epistemological stance concerning scientific knowledge is that knowledge is a passively acquired, faithful copy of the world, and all one must do is find it by looking in the right places. In order for students to move beyond this conception, we believe that they must have opportunities to become actively engaged in constructing and evaluating explanations for natural phenomena.

For the past several years, we have been developing and testing materials designed to help junior high school students come to understand the nature of scientific knowledge and inquiry more broadly conceived.

**The ETC Nature of Science Unit**

We have developed a month-long instructional unit which would replace the typical junior high unit on "the scientific method." Rather than focusing solely on how to carry out a controlled laboratory experiment, the Nature of Science Unit also emphasizes the continually building, intellectual nature of the inquiry process given specific phenomena amenable to such study by junior high school students. Unlike the typical unit which drills students on methodology using a number of unrelated phenomena, our Unit contains a few carefully selected and framed phenomena. Each phenomenon is studied in some depth for at least several days; the exception is the nature of yeast which is explored for two weeks. Such sustained investigation also contrasts with the usual approach to teaching about inquiry.

Portions of the unit were piloted with small groups of seventh, ninth and tenth grade students; in general, the results were encouraging and led us to begin to formulate the unit as a whole (Chomsky et al., 1985).

In the final, revised version of the unit, students' naive conceptions about a particular phenomenon motivate the inquiry process by provoking a question about an observed event. Students are asked to reflect on the the relationship between their own experience and knowledge about the phenomenon (or related phenomena) and their ideas about the problem. The choice of variables to be tested in an experiment are motivated by their ideas. Students are asked to reflect on the the thinking involved prior to, during, and after experimentation. Their growing knowledge about the phenomenon motivates further questions and continued investigation, and, in some cases, eventually leads to a change in their conception of the phenomenon.

We report here on two classroom studies using the Nature of Science Unit, first in its original form and later in a revised version (which was shortened due to time constraints).

**Overview of the Studies**

The studies reported here have two major research goals. First, we wish to probe junior high school students' initial understanding of the nature and purpose of science, and of the process of scientific inquiry. The most relevant assessment tool that currently exists is Klopfer and Carrier's (1970) Test of Understanding Science (TOUS), a test of what students know about science as a human endeavor and as a social institution. Included in what is probed by TOUS is knowledge of the vocabulary of science (e.g., hypothesis, law, theory) and the nature of scientific knowledge. TOUS probes whether students realize that scientific knowledge consists of ideas (i.e., concepts, laws, theories) about the natural world (i.e., its components, its inhabitants, what they are composed of, how they function and interact). Also probed is whether students understand that scientific knowledge is tentative, always subject to revision, and cumulative, building on the work of the past. We borrowed some of these items from TOUS as part of the multiple-choice, written pre/posttest instrument we devised. However, because such an instrument constrains students' responses, we felt that it might not reveal students' own ideas about the nature and purpose of the scientific process. Therefore, we also devised a clinical interview that allowed a deeper characterization of students' initial conceptualization of these aspects of the nature of science.

In addition, we explore whether it is possible to move students beyond their initial conceptions. The assumption underlying our approach to this problem is simple: one cannot teach students about scientific knowledge as intellectual construction without engaging them in the process of constructing scientific understanding. Our unit engages students in theory building, not only as a means of introducing students to the components of the scientific method, but more importantly to teach them about the nature of scientific inquiry and knowledge.

Study 1 reports the first field trial of our whole unit, comparing it with a well-developed, successful curriculum designed solely to teach process skills to seventh graders. In addition to the research goals stated above, at issue is whether the considerable material on the nature of scientific knowledge and inquiry in our unit leads to a decreased mastery of lessons on how to collect reliable data in a laboratory setting.

Observations and results from Study 1 led us to revise both the unit and student assessment measures. Study 2 reports a closer approximation to a real world trial, where a teacher (with minimal support) used the revised version of the unit in all of her seventh grade classes. While Study 2 did not compare our unit with any other treatment of the scientific method, the improved clinical interview allowed a clearer characterization of the students' initial conceptions about the nature of scientific inquiry, and a good assessment of what the students learned from the unit.

## STUDY 1: COMPARING THE NATURE OF SCIENCE UNIT AND A STANDARD UNIT ON THE SCIENTIFIC METHOD

### Overview

One purpose of Study 1 was to assess the feasibility of using the Nature of Science Unit with whole classes. Another purpose was to systematically compare its effectiveness with that of a standard unit on scientific method. A unit which is currently used in the Newton, Massachusetts school system and is known to successfully impart the process skills associated with laboratory experimentation served as the control unit in this study. Our questions were whether students would follow the more abstract metaconceptual points made in our unit, and whether they would master the points about collecting reliable data that are part of both units, but are the focus of much more practice in standard units, such as the one used in Newton.

**The Two Units**

The Control Unit: The Newton Unit on Scientific Method

A month-long unit on the scientific method begins seventh grade science in Newton. The goals of this unit center on the collection of reliable data. Students are taught the terminology of the experimental method: "hypothesis," "independent variable," "dependent variable," "controlled variables," "data," "measurement," and so forth. In the particular Newton classroom observed in our study, these ideas are introduced in the context of a classroom demonstration, in which the teacher puts cabbage juice in a variety of test tubes containing different clear liquids and the students observe many different color changes. Their task is to go home and experiment, attempting to discover under which conditions the different color changes occur. A class discussion of the resulting efforts is the occasion for introducing the components of the experimental method. The core of the Newton unit is a series of seven experimental work stations, each of which has an experimental question to answer (e.g., "What is the effect of liquid soap in water on the absorbancy of paper towels?"). The students work independently of the teacher, in groups of three to four. At each station, the group must formulate a hypothesis, design a test of it, carry out the experiment three times (for reliability), and draw conclusions from the data. In each case, the group must identify the independent and dependent variables, as well as the variables that are controlled. The purpose of these activities is to ensure that students understand the components of the scientific method.

The Experimental Unit: The ETC Nature of Science Unit

The goals of the experimental unit differ considerably from those of the Newton unit. To reiterate a point made earlier, rather than focusing on *how* one carries out an experiment, the Nature of Science Unit stresses *why* one might carry out an experiment, and the role an experiment plays in the development of an idea.

The unit begins with a week of lessons in which the experimental method is first introduced using two software packages: The King's Rule: Mathematics and Discovery (hereafter referred to as The King's Rule; Sunburst Communications) and The Scientific Method (Cygnus Software). The core of the Nature of Science Unit consists of two weeks of lessons on the nature of yeast and a week of lessons on linguistics. These two sets of lessons make somewhat different yet complementary points about the nature of science and inquiry. The yeast lessons exemplify the cumulative nature of science, the process by which a series of experiments can lead to an ever deeper understanding of a natural phenomenon. The linguistics lessons stress theory construction, especially the abstraction and justification of new theoretical entities, and the role of counterexamples in testing a theory. We briefly describe the lessons below.

*The Software Lessons*

The King's Rule introduces hypothesis formation and testing, emphasizing the relation between data and hypothesis testing and revision. Students are presented with a number triad and are asked to formulate a rule which generates the triad. The students test and revise multiple hypotheses by entering other triads into the computer, which indicates whether or not these triads follow the rule. The lessons we designed to accompany the software stress the importance of recording data, of accounting for all of the data, and of trying to falsify a hypothesis.

A version of The Scientific Method (revised so as to be more compatible with the goals and terminology of the unit) presents students with a wide variety of exercises which demonstrate various aspects of methodology, such as making careful observations, making and testing a hypothesis, manipulating and controlling variables, and drawing a conclusion. The main feature of the program is a simulation of an experiment on the chirp rate of crickets, which gives students an opportunity to evaluate the effect of a number of variables on chirp rate. Our lessons cover the logic of problem solving, the design of controlled experiments, hypothesis testing, and prediction.

*Yeast Lessons*

While the yeast lessons make the same the points about experimental methodology as the cabbage juice section of the Newton unit (e.g., the need for measurement, the need for controls), the lessons highlight the role of experimentation in broadening our understanding of a phenomenon. Students begin by discussing the question of what makes bread rise. By bringing the problem into a laboratory setting, they find that a mixture of sugar, warm water and yeast creates a gas. This explains the texture of bread, but raises a new question: why does this mixture produce a gas? Students learn that the answer to one question often leads to another question. Eventually, their inquiry leads to a fundamental question about the nature of yeast: Is yeast alive?

These lessons cover many points about the cumulative and fluid nature of scientific inquiry and knowledge. Students learn that in choosing variables and formulating hypotheses about a phenomenon, we are always guided by our current understanding. They also learn that the hypotheses worth testing are those that have the most theoretically important payoff; and that, despite the accumulation of evidence, experimental results may support but never prove a hypothesis.

*Linguistics Lessons*

In the linguistics lessons, students' implicit knowledge of their language is the object of inquiry. Students formulate rules to account for their intuitions about certain regularities in their language. They begin by trying to account for a syntactic phenomenon in speech, such as why one can contract "want to" to "wanna" in the sentence "I want to go," but not in the sentence, "Who do you want to go?" Students formulate and test hypotheses, and look for counterexamples to further test their generalizations. The explanation of this phenomenon involves a non-phonetic element which is interpreted to be between the "want" and the "to" in the second sentence, a theoretical entity that does much further work in syntactic theory. The role of this entity is also explored in the similar phenomenon of "is" contraction.

As this description indicates, the experimental unit covered most of the points in the Newton unit, plus many others about the nature of scientific inquiry and knowledge.

**Design of the Study**

The curriculum experiment was conducted at Day Junior High School in Newton in four seventh grade science classes that were normally taught by a single teacher. The experimental unit was one month long, and the control unit, several days shorter. Students in two of the classes (numbers 1 and 2) were considered by the school to be academically average or above average (hereafter referred to as "above average"), while students in the remaining classes (numbers 3 and 4) were considered to be average or below average (hereafter referred to as "below average").

The experimental group, which received the Nature of Science Unit, included 38 students from one above average class and one below average class (classes 2 and 4). The control group, which received the Newton unit, included 39 students from the remaining class in each level (classes 1 and 3). Table 1 illustrates this distribution of classes by curriculum and ability level.

| | | Curriculum | | |
|---|---|---|---|---|
| | | Control | Experimental | |
| Ability | Below Avg. | Class 1 N=2[illegible] | Class 2 N=16 | N=37 |
| | Above Avg. | Class 3 N=15 | Class 4 N=22 | N=39 |
| | | N=38 | N=38 | N=76 |

**Table 1. Design of Study 1**

The teaching was shared by the regular teacher and by a member of the Nature of Science group, each of whom taught one experimental and one control class every day. The regular teacher always taught the first control class, and the group member teacher always taught the first experimental class. Thus, each teacher had a chance to observe the lesson with which she was less familiar before she had to teach it.

Since the classes were on a rotating schedule, teacher differences were effectively cancelled; each teacher taught every class at least two times per week, and no more than three times per week.

**Assessment Measures**

All students (n=76) were given a group-administered written test during one class period before and after the intervention. Clinical interviews were also administered individually to 14 students (a representative sample from all classes) before and after the intervention. The interviews were conducted by members of the Nature of Science group, and lasted between twenty minutes and half an hour. All interviews were tape recorded and later, transcribed.

The Written Pre/Posttest

*Description*

We developed a 26-item written pre/posttest for use in the study. In preparing the pre/posttest, we reviewed a number of published tests of inquiry skills designed for the middle school grades, and found three tests to be most relevant to our work: Test on Understanding Science or TOUS (Klopfer & Carrier, 1970), Middle Grades Integrated Process Skills Test or MIPT (Cronin & Padilla, 1986), and Test of Enquiry Skills or TOES (Fraser, 1979). (For a detailed discussion of these and other tests, see Carey et.al., 1986.) Most of the items in the pre/posttest were adapted from these three tests. Additional items were developed by the Nature of Science group. Appendix 1 is a copy of the written test.

Our overall objective was to assess points covered in the experimental unit as well as points covered in the control (Newton) unit. The pre/posttest was thus designed to evaluate students' understanding of the nature of science and scientific inquiry, and also the logic of experimentation. The set of points emphasized in the experimental unit was assessed by test items on the nature of science and scientific inquiry (items 2, 4, 7, 12, 14, 15, 20, 21, 23) and the nature of theories (items 1, 3, 6, 13, 20, 22). Nature of science items included questions about the goal of science, the testing of ideas in science, and the notion that questions lead to new questions; while items about scientific theories tested the notion that theories are explanations for natural phenomena and are constructions of the mind. Another set of items assessed students' understanding of the components and logic of experimentation which was covered by both curricula but central to the control unit. These test items included: the identification and nature of hypotheses (items 7, 10, 21, 24, 25); experimental design (item 26); identification of variables (items 8, 9, 11, 16), the nature of controlling variables (items 17, 18, 19); and measurement (items 5, 18).

Within this set of test items, a number of sample problems required students to identify independent/dependent variables, hypotheses, experimental design or supporting evidence, while other items probed the nature or the purpose of these elements.

Preliminary versions of the pre/posttest were piloted with seventh and ninth graders to identify confusing or particularly difficult test items, and to determine whether the test could be completed in one class period.

*Scoring*

In scoring the test, each item was worth one point, except for items 5 and 24. Item 5 contained three true/false questions that were each worth one point. Item 24 required 12 responses; each correct response was worth a quarter of a point, for a total of three points. Thus, 30 points were possible on this test.

*Results*

Scores on the written pre/posttest were obtained from a total of 76 students: 38 from each of experimental and control conditions; 39 from the above average ability classes and 37 from the below average ability classes.

There was a modest and statistically significant overall improvement between the pretest scores and the posttest scores of 9.4 percent for all classes ($p<.0001$, paired comparison t-test, 1-tailed). Percent improvement was essentially the same for both experimental and control conditions (9.5 percent and 9.3 percent, respectively) and for the above average and below average levels (9.2 percent and 9.7 percent, respectively). These increases were significant at the $p<.0001$ level as well using a paired comparison t-test. However, comparing ability levels, there was an enormous difference among the mean scores on both the pretest and posttest by ability levels regardless of condition: above average students scored much higher (pre: 70.0 percent; post: 79.2 percent) than below average students (pre: 47.4 percent; post: 57.1 percent). The difference between the two ability levels on the pretest and posttest were significant at $p<.0001$ (pretest: $t=8.2$, d.f.=74, $p<.0001$; posttest: $t=7.8$, d.f.=74, $p<.0001$). See Figure 1.

Using an ANOVA statistical analysis, the interaction between condition and ability level approached significance ($F=3.38$, d.f.=1, $p<.10$). In the experimental classes, the above average experimental class improved more on the posttest than did the below average class. In the control classes, the below average control class improved more on the posttest than did the above average class. These findings hint that the experimental unit may be better for students of above average ability and that the control unit may be better for students of below average ability.

As mentioned above, the pre/posttest was not designed specifically for either the experimental or the control unit. Therefore, we examined test performance according to which items were emphasized by each of the units. We found no significant difference in the degree of improvement between the experimental group and the control group on either the set of items about the nature of science or the set

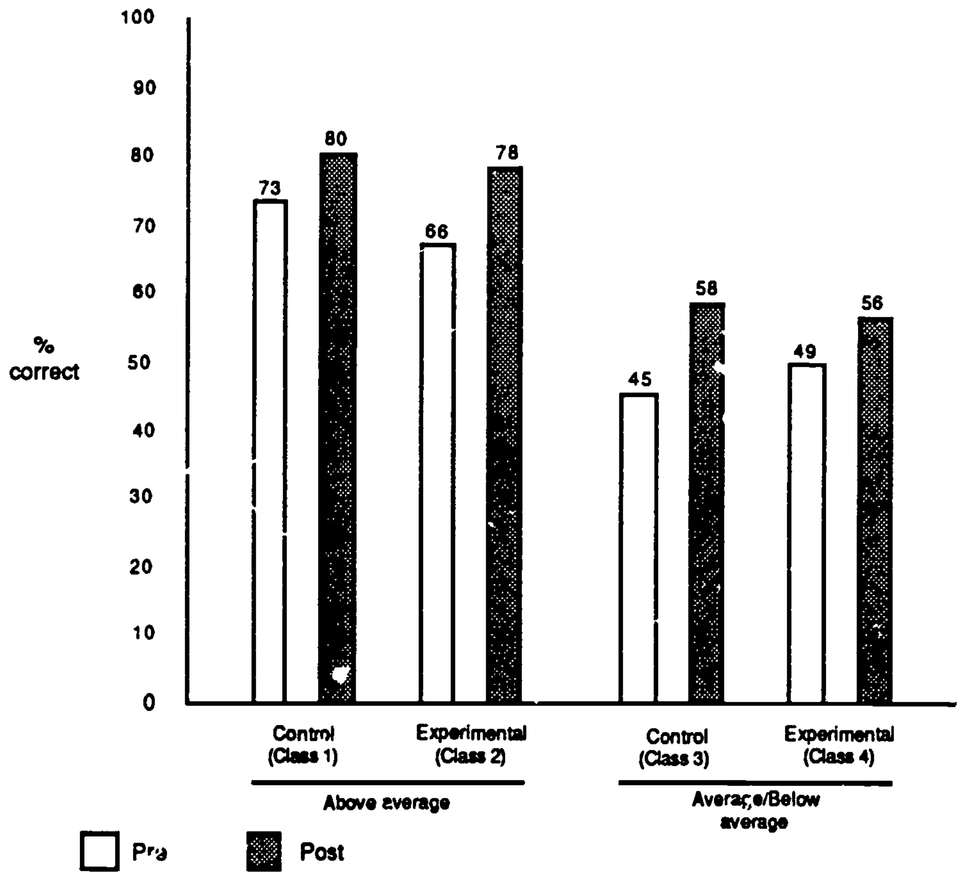



**Figure 1. Pretest and Posttest Mean Scores by Condition and Ability Level**

of items testing the components and logic of experimentation. For the most part, the experimental and control classes improved equally on both types of items. Significantly, the experimental class did not learn *less* about collecting reliable data than did the control class. Equally significantly, the control class improved as much as did the experimental class on items designed to tap knowledge of the nature of science.

## The Clinical Interview

### *Description*

The half-hour clinical interview was designed to characterize an individual student's initial understanding of the nature of science and scientific inquiry before and after the unit. While the interview did not directly address the content of either the experimental or the control unit, it allowed us to probe a student's ideas about the nature and purpose of theories, hypotheses, experiments and science in general through open-ended questions. Appendix 2 is a copy of the clinical interview.

The interview consisted of the following seven sections: the nature/purpose of science and theories (items 1a-d, 2a-c); the origin of theories (items 3a-c); process of theory development (item 4); results and evaluation (items 5, 10); the role of guiding ideas and questions (items 6, 8); the nature/purpose of experiments (item 7a-b); and the nature/purpose of a hypothesis (item 9a-b).

### *Analysis of the Interview*

The process of analyzing the interviews was iterative. The interviews were transcribed and read impressionistically at first, and a range of answers for each question was established. Based on this range, a loose coding system was developed for each set of related questions. The coding system was modified in form and in content after each round of scoring, and again after the code for the clinical interviews from Study 2 was completed.

As the scoring criteria became sharper, two major problems with the design of the interview began to appear. The first problem was the absence of a certain type of question. A crucial part of the definition of any element of the process of scientific inquiry (i.e., the scientist's ideas, experiments, results/data) is the relationship between that element and the other elements. For example, a complete understanding of a hypothesis includes the notion that it may be tested in an experiment. Yet, while each question in the interview explicitly addressed a particular element in the process of scientific inquiry, none addressed the relationships among these elements. As a result, many of the students did not talk about these relationships in their answers, and it was difficult to gauge how deep their understanding of the inquiry process really was.

The second problem was the lack of follow-up questions. Students often responded to questions with words like "research," "test," "experiment," or "prove." While these words certainly sound meaningful, it is often difficult to tell just what they mean to a particular student. A student could use the word "experiment" to mean anything from mixing chemicals together in a beaker in order to see what happens to doing a carefully controlled test of an idea. The fact that students were not asked to explain themselves when they used such "big" words also made it difficult to assess their understanding.

Because it was not closely tied to either the Newton unit or the Nature of Science Unit, the interview was not a valid measure of the effectiveness of either curriculum. In addition, the sample of students who received the interview was too small to reliably represent the population of each class. And, as we have already mentioned, the questions did not elicit students' general concepts as thoroughly as we had hoped they would. The interview did, however, through its shortcomings, guide us in the design of the clinical interview for Study 2, which was administered while we were still developing a coding scheme for the interview of Study 1.

### *The Coding Scheme*

The coding scheme for the interview included three general levels of understanding (excluding Level 0, in which there is no sense of the purpose or activities of science). In General Level 1, the student makes no clear distinction between ideas and experiments. Thus, he or she believes that the motivation for any activity is in the characterization of the activity itself, and does not relate the activity to the construction or verification of ideas. The point of an activity may be to discover a fact, invent something, or find out what happens when you do "X." Scientists' theories are about (or equivalent to) these activities. In General Level 2, there is a clear distinction between ideas and experiments. The motivation for activity is the verification or exploration of an idea; the purpose of doing an experiment is to test an idea to see if it's right. In General Level 3, the motivation for activity is again verification, but added to the distinction between ideas and experiments is an understanding of the relationship between results (especially unexpected ones) and the idea being tested. This includes an appreciation that ideas may be changed when the evidence demands it.

As described above, the interview was divided into sections, each of which contained one or more items. Within a section, closely related questions and probes were usually scored as one "item." For example, questions 2a through 2c (Have you ever heard of the word "theory?" What are theories all about? What does a theory do?) were scored as one of the two items included in The Nature/Purpose of Science and Theories section. Each item was scored according to a version of the general coding levels adapted for the particular section in which it appeared.

The general coding levels and their adaptation for scoring each section of related interview items are described in detail in the discussion of the Study 2 clinical interview.

*Results*

Given the small number of subjects (n=14), the quantitative analysis of the interviews was pursued merely to gain a rough idea of the level of response and any strong differences between groups. For a detailed description of this analysis, see Appendix 3.

Overall, the mean score per-item increased from .85 to 1.22 and this increase was significant ($p<.01$, Wilcoxon signed ranks test, 1-tailed). The result shows that all students made some gain from a very limited understanding of the purpose and activities of science. Although some of the group differences were significant, in general they were not strong. We believe, however, as noted above, that the interview was not an effective measure of the important aspects of either curriculum.

**Critique of the Experimental Unit**

The following critique of the lessons in the experimental unit is based on observations made by the teachers and group members during the study. These observations led us to revise the unit for Study 2; revisions are also briefly noted below.

Software: The King's Rule and The Scientific Method

In The King's Rule, students formulated and tested hypotheses about the rule underlying the generation of various number triads. In The Scientific Method, students solved a series of unrelated problems that emphasized different parts of the scientific method. The software was somewhat successful in teaching students how to formulate and test a hypothesis, but much less successful in teaching them how to interpret the results of a test with respect to a hypothesis or how to revise the hypothesis based on the results. It taught nothing about science as the development of ideas.

In neither The King's Rule nor The Scientific Method were problems or questions motivated by the students' own observation of a phenomenon; they were simply given by the software. The software never required students to reflect on the role their own ideas and intuitions played in solving those problems. Students did not evaluate and develop their ideas in a way that led to any kind of new understanding. Nor did they choose variables that they thought were relevant to the solution of a problem. Instead, the variables and solutions for the problems were quick, absolute, and usu[illegible] easily gotten from the computer. In general, then, the software reinforced the i[illegible] that science is about solving an assigned problem by finding the "right" answer in some external source (here, the computer) and moving on to something else, rather than about constructing ideas.

The lessons we designed to accompany the software were difficult to implement. Although the teacher sometimes interrupted the computer work to make points about the process of inquiry to the class as a whole, this was difficult to do because the students worked at varying speeds, and thus were often working on different problems.

For these reasons, The King's Rule and The Scientific Method software were not included in the revised unit.

Yeast Lessons

In the yeast lessons, students considered the difference between frozen bread dough and a piece of bread, and explored the question of what makes bread dough rise. They worked through several cycles of forming and testing hypotheses in controlled experiments, and built a theory about the phenomenon that includes the notion that yeast is alive.

We encountered some serious problems with the yeast lessons, the solutions to which guided us in our revision of the lesson plans for Study 2. In general, the lessons were overambitious in terms of both the material to be covered and the logistical requirements. The teacher was required to introduce a number of new points about the nature of science each day. As a result, some of the points were never made explicitly, and many of them did not receive enough articulation and reiteration to make them clear to the students. Also, the lessons required more equipment and laboratory preparation than a single teacher could effectively handle (fortunately, group members were present to help at crucial moments). The students themselves were unfamiliar with some of the equipment ("How do you use a thermometer?"), making class management all the more difficult.

For Study 2, some of the more complicated labs were rewritten as classroom demonstrations, allowing more time for class discussions and review of the theory building process. In addition, selected points about the nature of science were clarified and reiterated throughout the revised yeast lessons.

Another observation made by one of the teachers was that the yeast (and software) lessons did not contain enough written exercises. For Study 2, more homework and lab sheets were created so that students would be required to think carefully about their ideas and experiments between lessons.

### Linguistics Lessons

In the linguistics lessons, students observed several language phenomena, and developed explanations for them by formulating, testing, and revising hypotheses. Some of the students were skeptical about the point of the lessons ("Why are we doing English in science class?!"). Others found the process of puzzling over data which they had generated themselves to be challenging and satisfying.

These lessons required more teacher-training than the other subunits. In addition to being familiar with the lessons themselves, the teachers needed to be able to deal with unexpected hypotheses presented by the students. The teachers had to treat these hypotheses the same way as they treated the hypotheses predicted in the lesson plans, by testing them and looking for counterexamples. When they were not able to do this, the focus ended up on the question of whether or not a student's hypothesis was "right," rather than on the process of inquiry itself.

The linguistics lessons were revised to contain contingency plans for dealing with a wider variety of student hypotheses.

### Some General Comments About the Unit

Beyond the relatively straightforward and easily solved logistical problems of the experimental unit, the above critique implies another, perhaps more important reason for why our unit sometimes failed to make the intended points about the nature and process of scientific inquiry.

Some students were able to tease an understanding of the process of scientific inquiry from the content in which it was embedded. When asked informally what they had learned from the unit, they answered with statements like, "how to make a hypothesis and test it," "how to do an experiment," "when you solve one problem, it leads to a whole new problem," and "we learned about a natural phenomenon." In this sense, the experimental unit was successful, since the students learned many of the same things that they would have learned in the standard unit.

For other students, even the methodological points got lost in the content. These students said they had learned things like "how to use a computer," "how to look for patterns in numbers," "sugar and water and yeast bubble when you mix them," "how to make bread rise," and "how to contract words."

Even those students who did understand the points about methodology that were also included in the standard unit did not seem to appreciate the broader purpose of the methodology; the investigation of natural phenomena and the construction of ideas. We suspect that the absence of a thorough understanding of the way the scientific method leads to a deeper understanding of the world, as well as the emphasis on content in the minds of many of the students, can be attributed, at least in part, to the implicit way in which many of the process-of-inquiry points were made in the lessons, and to insufficient review of these points. Keeping in mind the difficulty students had abstracting the implicit ideas from the content in which they were embedded, we articulated the points more explicitly and repeatedly in the revised lesson plans. We attempted to make the motivation for each part of the process of inquiry more overt, and to continually relate the process to the bigger picture -- the problem being investigated, and the idea being developed.

**Conclusions to Study 1**

Study 1 had two main goals. First, and most simply, we wanted to see whether the experimental unit is managable in the whole classroom. (The various lessons had previously been piloted only with small groups of students.) There is no doubt that the unit is teachable, at least under the special conditions of this study where there were always two teachers present in the classroom. In part, our revision of the lesson plans for Study 2 was intended to make the lessons managable by a single teacher.

The second purpose of the study was to learn whether the experimental students would follow the more abstract points about the nature of science and scientific inquiry contained in this unit <u>and</u> also master the skills of experimentations that were the focus of the Newton unit. This is important because the experimental unit is much more ambitious in its goals. A serious possibility was that students would become overloaded with all they had to manage -- serious new science content (about yeast and linguistics), points about the experimental method, and points about theory building. This did not happen. The most salient aspect of the results is the comparability of the experimental and control groups in terms of improvement on both the written test and the clinical interview.

There was a hint that the experimental unit, in its Study 1 incarnation, was more effective for the above average student, and the control unit was more effective for the below average student (although the interaction between condition and ability level on improvement on the written test did not reach the .05 level of significance using an ANOVA.). One possible explanation is that the control unit had many repetitive exercises, with much more writing required of the students. As discussed above, our unit has been revised to remedy this situation.

Three aspects of the results require comment. First, the improvement between pretest and posttest, while significant, was not dramatic. The main reason for this is that the test was not geared particularly to either curriculum. Second, there was no difference between the experimental and control goups in their improvement on the written test and clinical interview items that covered points made only in the experimental unit. There are several reasons for this. The same teachers taught both units, and the Newton lessons were inevitably influenced by the experimental unit. Also, although there were some sections in the clinical interview on which only the experimental group improved, a fine-grained comparison between the groups is not meaningful because of the few students involved. Finally, because of the limitations of the clinical interview, this study did not allow as full a characterization of the students' initial understanding of the nature of science as we would have liked. These issues were addressed in Study 2.

## STUDY 2: THE REVISED UNIT
## A CLOSER APPROXIMATION TO A REAL WORLD TRIAL

### Overview

Study 2 had two main goals. The first goal was to assess changes in students' understanding of the nature of science more sensitively. In order to do this, the clinical interview was revised to allow a more adequate characterization of seventh graders' understanding of the nature of science, and to allow us to probe more specifically points about the nature of science and scientific inquiry made in our curriculum. It was administered to over one-third of the students, providing a good assessment of the effectiveness of the curricular intervention. The second goal of Study 2 was the to improve the unit and to observe it used in a typical classroom situation. Therefore, we prepared revised lesson plans that would allow a teacher to teach the unit on his or her own, without a member of our group acting as a model or co-teacher. In contrast to Study 1, Study 2 did not compare our unit with any other treatment of the scientific method.

The Nature of Science Unit was revised for Study 2 in the light of our experience in Study 1. One main lesson from Study 1 was that the points we were trying to impart about the nature of science and scientific inquriry were too diffuse and poorly articulated. Not enough time was alloted for the explicit articulation and discussion of these points. We found that it is all too easy to concentrate on the content of the science being taught, leaving the metaconceptual points about the nature of science implicit. In the revised unit, these points are explicitly dealt with, particularly in lessons designed specifically for their discussion. Written exercises were also devised to help students think about their ideas and experiments between lessons.

In addition to these general modifications to the lessons, there were changes in the structure of the month-long unit. These included: (1) the addition of an introductory lesson and a wrap-up lesson, (2) replacement of the computer software lessons with lessons using an ETC-produced interactive videodisc, and (3) revisions to the yeast lessons, making them manageable by one unassisted teacher. Given constraints that restricted teaching time to three weeks instead of one month, the linguistics portion of the unit was not used in Study 2. (Another classroom study using a revised and expanded version of the linguistics lessons was carried out in the fall of 1987 and will be reported on in future work.)

**Description of the Revised Unit**

Introductory Lesson

The introductory lesson gives students a chance to speculate about a "martian object" that the teacher displays before the class. The students come up with questions about the nature of the object, and eventually focus on the question of whether or not it is alive. They come up with a list of attributes of 'aliveness', and discuss ways they could test to see which (if any) of these attributes the object possesses. The students reflect on the fact that their ideas about aliveness come from their own experience, and reside in their own minds.

Embedded in the lesson are the points that:

(1) The basis for scientific inquiry is mental work; and

(2) Experiments are tests of ideas

The Videodisc Lessons: Seeing the Unseen

In 1985-86, the ETC Videodisc Project produced Seeing the Unseen, a four-segment interactive videodisc designed to teach middle school students about the nature of science. Its use by individual students, by pairs of students, and by whole classrooms had been explored (Storey and Mellin, 1987), and it had been concluded that its effectiveness would be enhanced if the points it contained about the process of inquiry were explicated by the teacher and discussed by the class. Study 2 allowed us to explore this possibility. Two segments of the videodisc engaged students in pieces of theory building, so these were chosen for for the revised Nature of Science Unit.

*Animal Mimicry Lessons*

In the animal mimicry lessons, which are designed to support the videodisc segment called "What Disguises Do Animals Use," students use their initial ideas about the basic needs of animals to construct their own categories of different kinds of animal disguises. On the first day, after viewing a video clip showing animals which disguise themselves in various ways, students are asked to generate their ideas about the needs of animals and to begin applying what they know toward understanding the functions of different kinds of disguise.

With these ideas in mind, they then view six other animal disguise clips with the task of devising categories that they think are best. On the second day, they discuss the categories they have devised and check them against additional data.

The teacher points out that the process they have gone through is similar to the way a scientist might go about categorizing the same data.

The points made in this lesson include:

(1) Ideas, including categories or systems of classification, help us make sense of the world; and

(2) Categories are mental constructs that are developed by observing and thinking.

*The Black Box Problem*

In the lessons for the Black Box Problem, students try to figure out the regular, geometric shape of a three-dimensional object in a closed box. First, students are allowed to brainstorm about methods which may help them figure out what the shape of the object is. Then they view an interactive videodisc segment called "How do Scientists Study Things They Can't See?" which shows Linus Pauling verbalizing his thought process while he works on the black box problem. The videodisc is stopped at certain points to discuss Pauling's systematic method for hypothesis testing.

On the second day, students are given sealed boxes with objects inside , and must work on the problem themselves. Students must defend their hypotheses about the shape of the object on the basis of tests they perform on the black box.

The lessons make the points that:

(1) A useful way of solving a problem is to break it down into smaller, more investigable parts;

(2) An experiment is a test, the results of which will either support or disprove one's hypothesis;

(3) One makes a hypothesis before doing a test, and interprets the results of the test with reference to the hypothesis.

(4) Scientific inquiry can proceed even when the object under investigation cannot be seen or touched; and

(5) We may never know the "right" answer, only best approximations.

Yeast Lessons

The yeast lessons include a series of brainstorming sessions, experiments, and demonstrations, during which students have the chance to develop an ever-deepening understanding of a natural phenomenon. The exploration begins by observing and discussing the difference between a piece of bread and a piece of unrisen bread dough, which leads to the question, "What makes bread rise?" Students reflect on what they already know about bread and are encouraged to continue relating their existing knowledge to their investigation. They first think about the ingredients that are used in bread making and the fact the dough requires a warm place in order to rise. In pursuing their question, they begin experimenting with laboratory versions of the phenomenon, testing the ingredients that seem relevant.

After observing that a flask containing yeast, sugar, flour, salt and water creates bubbles (a gas), the students conclude that the production of a gas makes bread rise. This answer just leads to another question: What produces the gas? Students then perform their own experiments to determine which ingredients are necessary for the mixture to bubble. Their initial experiments are often unsystematic, so the teacher discusses with the class the importance of controlling variables in experiments. Pooling their results as a class, they then repeat their experiments as a controlled experiment.

Once the "what" question has been answered, the focus becomes "Why do the ingredients in the mixture produce a gas?" After brainstorming about what they know about each of the necessary ingredients, students formulate hypotheses about what is happening in the flask. In the course of the discussion, two competing hypotheses are raised: (1) The bubbles are caused by some kind of chemical reaction, and (2) Yeast is alive, the yeast eats the sugar, and the gas is a product of metabolism. Thus, the students are led to a fundamental question about the nature of yeast: Is the brown granular substance a chemical, or is it a living thing? Students are encouraged to think about experiments they might do to decide between these hypotheses. As a class, they do experiments to test what kind of gas is produced by the reaction, what happens to yeast that is subjected to extreme temperatures, and what happens when frozen yeast is thawed. Each time they assess the results in light of the two proposed hypotheses. The results of these experiments seem to support the hypothesis that yeast is alive. At the end, they use what they have learned to explain various real world phenomena about the behavior of bread dough.

The yeast lessons make a number of points:

(1) Scientists solve problems through the interplay of the ideas they have, the experiments they choose to test those ideas, and new ideas that arise from the results of the experiments;

(2) One of the goals of science is to arrive at explanatory understandings of natural phenomena;

(3) The answer to one question often leads to another question; and

(4) In exploring a phenomenon, scientists may be forced to challenge some of their basic notions about the way things are.

Wrap-up Lesson

The wrap-up lesson concludes the unit with a general discussion about the interplay of thought and experimentation in science. It reviews the methods of scientific inquiry the students have used, making specific references to previous lessons.

Some overall items covered include:

(1) The point of exploring problems is not just to solve the problem, but rather to develop a deeper understanding of things in nature;

(2) Scientific inquiry involves an interplay of thinking and experimenting;

(3) Questions change during the course of scientific inquiry; and

(4) Results of investigations often challenge our basic notions about the way things are.

**Design of the Study**

The three-week long study of the Nature of Science Unit was conducted at the Hosmer East School in Watertown, Massachusetts. The unit was taught to 76 students in five mixed ability seventh grade science classes by a single teacher who was their usual classroom teacher.

**Assessment Measures**

Assessment measures included a written test that was administered to all students before and after the intervention. Clinical interviews were also administered individually to randomly selected students from each of the five classes before and after the intervention (Total n=27). The interviews were conducted by members of the Nature of Science Group, and lasted between twenty minutes and half an hour. All the interviews were tape recorded.

Daily observations were made by a member of the research team. Teacher reactions and comments were elicited through informal interview. A survey of student reactions was taken at the end of the intervention.

The Written Pre/Posttest

*Description*

The 24-item written test was designed to assess the extent to which students learned specific points that were covered in the Nature of Science Unit. The pre/posttest evaluated students' understanding of the nature of science and scientific inquiry; the role and nature of hypotheses, experiments, and ideas; and the logic and purpose of systematic aspects of experimental design. Appendix 4 is a copy of the written multiple choice test.

While it contained many of the same items as did the written test of Study 1, the Study 2 pre/posttest was reworked to capture more accurately the important points of the Nature of Science Unit. Items that were not directly related to the unit were removed and a number of new, more specific items were added. Since the unit itself was changed and new points made in the Study 2 unit, additional test items to assess the new points were created.

Test items were grouped into the following sections so as to analyze more precisely the effect of the Nature of Science Unit on particular areas.

*Nature/Purpose of Science and Scientific Ideas.* The Nature of Science section includes questions about the goal of science, the nature of scientific theories, and the testability of scientific questions (items 1, 8, 9 and 16).

*Development of Ideas.* The questions in the section on the Development of Ideas deal with the notions that questions change during scientific inquiry, that questions lead to new questions, that ideas are developed through making predictions and testing one's ideas, and that evidence sometimes cause scientists to change basic ideas about the world (items 2, 7, 10, 19).

*Identification of Hypothesis.* This section consisted of one item in which students were to identify the hypothesis being tested in a sample situation (item 4).

*Nature/Purpose of Experiments.* In addition to items about the purpose of doing experiments, in several example situtations, students were asked to identify controlled variables and indicate the purpose for controlling variables and measuring variables (items 5, 12. 13, 14, 20).

*Guiding Ideas and Questions.* Some items in this section focused on ideas as a guide for doing experimentation and for deciding what variables to test in an experiment. It also included an item where students were asked to indicate the sequence of what one would do in an experiment (items 6, 11, 17).

*Results and Evaluation.* This section contained items about conflicting evidence and the purpose of making predictions in experimentation. A number of items in this section involved sample situations where students were to indicate whether they thought the particular information gathered supported or did not support a given hypothesis (items 3, 18, 21-24).

### *Scoring*

In scoring the test, each item was worth one point, except for item 17 where there were two points possible. In item 17, one point was credited if the student correctly identified the first step in the sequence; an additional point was credited if the student ordered the remaining steps correctly. Thus, 25 points were possible on the test.

### *Overall Results*

Scores on the written pre/posttest were obtained from a total of 73 students. Fifty-nine students completed both the pretest and posttest and could be included in the pre/posttest comparison. The results are reported in the form of the percent of items correct for all students in that sample. Table 2 presents the overall results and the results by section.

Overall, there was a small, but significant difference on a 1-tailed paired comparison t-test between the average percent correct on the pretest and the average percent correct on the posttest, an overall improvement of 5.9 percentage points ($p<.0005$, $df=58$). The pretest mean of 68.5 percent correct increased to a posttest mean of 74.4 percent.

| | PRE | POST | DIFFERENCES |
|---|---|---|---|
| NATURE/PURPOSE OF SCIENCE | 66.1 | 79.7 | 13.6 *** |
| DEVELOPMENT OF IDEAS | 72.0 | 82.2 | 10.2 ** |
| IDENTIFICATION OF HYPOTHESIS | 67.8 | 81.4 | 13.6 * |
| NATURE OF EXPERIMENTS | 61.4 | 69.8 | 8.5 * |
| GUIDING IDEAS | 70.3 | 70.8 | 0.4 |
| RESULTS/ EVALUATION | 72.6 | 72.6 | 0.0 |
| OVERALL | 68.7 | 74.6 | 5.9 *** |

* $p < .05$
** $p < .005$
*** $p < .0005$

**Table 2. Mean Percent Correct on Study 2 Written Pre/Posttest (n=59)**

*Results by Section*

In both the Nature of Science section and the Development of Ideas section, there was a significant increase in scores using a 1-tailed paired comparison t-test ($p<.0005$ and $p<.005$ respectively). The Nature of Science section increased from 66.1 percent correct to 79.7 percent, a pre to post difference of 13.6 percentage points. In the Development of Ideas, there was a 10.2 percentage point increase, from 72 percent to 82.2 percent correct.

The Identification of Hypothesis section and the Nature/Purpose of Experiments section both showed a statistically significant increase on a 1-tailed paired comparison t-test (both, $p< .05$). In the Identification of Hypothesis section, the pretest score of 67.8 percent increased to 81.4 percent on the posttest, a difference of 13.6 percentage points. The Nature/Purpose of Experiments section scores raised 8.5 percentage points, from 61.4 percent to 69.8 percent.

In the Guiding Ideas and Questions section and the Results and Evaluation section, there was no significant increase.

*Discussion*

The most significant increases in test scores occurred in the sections on the Nature of Science and the Development of Ideas. This is not surprising given that the questions in these two sections tested students' understanding of points that were stressed in the unit.

One of the main goals of our unit -- particularly of the yeast lessons -- was to teach about the nature and purpose of doing science, principally that the aim of science is to build explanations of natural phenomena. This point was reiterated throughout the unit. In test items about the purpose of science and scientific theories, there was a pretest to posttest decrease in the number of students who responded that science consists of uncovering facts about nature, and a corresponding increase in the number of students who responded that the aim of science is to develop the best explanation of things in nature. This is precisely the shift we would anticipate as a result of the unit.

In the Development of Ideas section, several relevant points -- for example, that questions lead to new questions and that basic ideas about the world are often challenged -- were not only discussed but also exemplified in the students' progress through the yeast lessons. Here too, the shift in student answers is interesting. Students moved away from the response that scientists almost always use experimental results as solutions for making inventions and toward a response that the experimental results lead to new questions. Similarly, their notion that scientists develop their ideas simply by trying things and seeing what happens gave way to the notion that progress occurs through making predictions based on an idea and testing those predictions.

The Identification of Hypothesis section and the Nature/Purpose of Experiments section also showed significant improvement from pretest to posttest. As in the previous two sections, there was a shift in students' understanding of the purpose of doing experiments. Responses which mentioned seeking inventions or simply seeing what happens when things are mixed together as the reason for doing experiments decreased, while responses which attributed exploration and testing of the scientists' ideas as the purpose of experiments increased. Again, the shift is in the direction hoped for as a result of the unit. The other points tested in this section, including what a hypothesis is and the rationale for controlling variables, were covered in both the black box lessons and the yeast lessons.

There was little change in the overall pretest to posttest scores in the Guiding Ideas and Questions section. However, upon closer examination of the scores by test item, it was found that an increase in correct responses for one of the sequencing of experimental steps items (item 17) was offset by a decrease in correct responses to an item about how one decides what variables to test (item 11). Scores for the other two items in the section remained the same. Hence, there actually was some improvement in this section, except on the particular item about choosing variables. A possible explanation for the drop in correct responses for this item might be as follows. In the classroom, there was explicit discussion of how one chooses variables and students participated in selecting appropriate variables; however, when it came time to do the yeast experiments, the teacher assigned which variables each group of students would test. It is possible that the assigning of variables led students to believe that variables are given in the problem and that every variable that can be tested should be tested (the two incorrect responses frequently chosen), not keeping in mind that the basis for choosing the original list of variables was their ideas about what is relevant. Further development of the lesson plans should take this problem into account and attempt to find ways to reduce this misunderstanding.

There were no changes from the pretest to posttest on the Results and Evaluation section. This may be a consequence of the test items. Nearly half of the section score came from a task involving identification of supporting evidence given a hypothesis (items 21-24). Administration of items of this type in Study 1 showed that students were bored by doing the same exact task on the pre/posttest, and thus some did more poorly on the posttest. While we administered different versions of this task on the Study 2 pre/posttest, the simlarity of the items may still affect student performance.

Although we still believe the clinical interview to be a more sensitive instrument when it comes to measuring the effects of the Nature of Science Unit, we are encouraged by the significant overall improvement in written test scores after presentation of the unit.

## The Clinical Interview

### *Description*

The Study 2 interview was closely tied to the Nature of Science Unit revised for this study. It posed new questions about the relationships between ideas, experiments, and results, and also included follow-up questions. In addition, it was administered to a much larger sample of students than was the Study 1 interview. As a result, the Study 2 interview was much more effective at eliciting complete answers from a representative range of students.

The revised clinical interview probed students' understanding of the following: (1) the nature and purpose of science; (2) the main elements of scientific work, including ideas, experiments, and results/data; and (3) the relationships among these elements, and motivations for the relationships. As in Study 1, the purpose of the interview was both to characterize the students' initial understanding, and to chart any movement that may have resulted from the teaching intervention. Appendix 5 is a copy of the revised clinical interview.

### *Analysis of the Interview*

The responses to each part of the interview were coded into categories that reflected three general levels of understanding, which are described below (excluding Level 0, in which responses revealed no sense of the nature of science). The coding scheme was developed on protocols taken from students who received only the pretest or the posttest, and whose data were not included in the final analysis. The tapes were not transcribed; each interview was coded by at least two people, who were unaware of whether it was a pre-interview or post-interview. Interscorer reliability was modest (74 percent agreement; disagreeme ... ually always involved only one level, such as one scorer judging a response t ... rth a 1, while another judged it a 2). Scoring disagreements were resolved by dis ... . When students responded "I don't know," no score was given; Level 0 responses were actual misconceptions or irrelevant answers (e.g.,"an experiment is when you mix two things together," or, "a hypothesis is a kind of snake.")

Questions related to a single theme were grouped into sections. For each section, every student received a mean score, and the overall mean for the section was calculated from these scores. In addition, the highest score a student attained in a section was noted, and the overall mean high score was calculated for each section. Because the the point of the interview was to elicit students' conceptions and misconceptions about the nature of scientific inquiry, rather than to test their vocabulary or make an inventory of the things they had no ideas about, questions on which the students received no score were not included in the calculations of any of the means. This may have slightly raised the means on the pre-interview, but did not have an appreciable effect on the scores of the individual sections, since nearly all of the "I don't know's" appeared in the pre-interview Hypothesis section.

*The Coding Scheme*

The students' ideas about the nature of science ranged from a notion that doing science means discovering facts and making inventions, to an understanding that science is about constructing explanations for natural phenomena. The general coding levels reflect this range along several dimensions. The first dimension is the degree to which ideas, experiments, and results are defined and differentiated from one another. The next is the degree to which the relationships among these elements are articulated and understood. Together, these dimensions define the motivation for a scientist's activities, which ranges from a simple carrying-out of a plan to a guided exploration of a phenomenon or construction of an idea.

In General Level 1, the students make no clear distinction between ideas and activities, especially experiments. A scientist "tries it to see if it works." The nature of "it" remains unspecified or ambiguous; "it" could be an idea, a thing, an invention, or an experiment. The motivation for an activity is located in the characterization of the activity itself, rather than in the larger context of the construction of ideas. Scientists discover facts, invent things, or find out what happens when they do "X," and the goal of science is equivalent to the activity itself.

In General Level 2, students make a clear distinction between ideas and experiments. The motivation for activity is verification or exploration of an idea, and more specifically, the purpose of doing an experiment is to test an idea to see if it is right. There is an understanding that an idea or phenomenon may be operationalized and explored. The goal of science is to find out how things work.

In General Level 3, as in Level 2, students make a clear distinction between ideas and experiments, and understand the motivation for activity as verification or exploration. Added to this, however, is an appreciation of the relationship between the results of an experiment (especially unexpected ones) and the idea being tested. An idea is evaluated in terms of the results of a test, and may be changed or developed in accordance with these results. Thus, Level 3 understanding recognizes the cyclic, cumulative nature of science, and identifies the goal of science as the construction of ever-deeper explanations of the natural world.

The general levels are a guideline for all the sections. The specific coding levels will be presented in the discussion of each section.

*Results by Section*

For the purpose of scoring, the interview was broken into six sections: (1) the nature/purpose of science and scientific ideas; (2) the nature of a hypothesis; (3) the nature/purpose of an experiment; (4) guiding ideas and questions; (5) results and evaluation, and (6) relationships. We will characterize the questions and coding levels for each section and describe typical student answers. The overall mean scores and mean high scores for each section are shown in Table 3. The number of students scoring equal to or above specific levels overall and by section is shown in Table 4 (next page).

| | Mean | | | | Mean High | | | |
|---|---|---|---|---|---|---|---|---|
| | Pre | Post | Diff. | Sig. | Pre | Post | Diff. | Sig. |
| Nature/Purpose of Science/ Scientific Ideas | 1.09 | 1.28 | +.19 | p<.03 | 1.52 | 1.74 | +.22 | NS |
| Nature of a Hypothesis | 1.0 | 1.37 | +.37 | NA | 1.0 | 1.37 | +.37 | NA |
| Nature of an Experiment | 1.0 | 1.52 | +.52 | p<.002 | 1.0 | 1.52 | +.52 | p<.002 |
| Guiding Ideas & Questions | .65 | 1.45 | +.80 | p<.001 | 1.19 | 2.22 | +1.03 | p<.001 |
| Results & Evaluation | 1.06 | 1.8 | +.74 | p<.001 | 1.41 | 2.19 | +.1.78 | p<.001 |
| Relationships | .91 | 1.69 | +.78 | p<.001 | .93 | 2.3 | +1.37 | p<.001 |
| Overall Score | 1.0 | 1.55 | +.55 | p<.001 | 1.93 | 2.63 | +.7 | p<.001 |

**Table 3. Mean scores and mean high scores by section for the Study 2 clinical interview (n= 27)**

*(1) The Nature/Purpose of Science and Scientific Ideas.* The questions included in this section are: What do you think science is all about? What do you think is the goal of science? Which is statement better describes the goal of science: "The goal of science is to discover new things in the world and the universe" or "The goal of science is to build a better understanding of the world around us"? Why? What kinds of ideas do scientists have? What are their ideas about?

Level 1 answers focus on the activities themselves. The point of science is to find out or discover facts and answers, or to invent cures or contraptions, and scientists have ideas about how to carry out these activities. In typical student answers, the goal of science is "to discover new things," "to find new cures for diseases," "to discover all the facts there are about the world." Scientists have ideas about how to discover or invent things, about "finding out about the stars and nature," about "teaching us and doing experiments."

| Overall | | | Nature/Purpose of Science | | | Nature of a Hypothesis | | |
|---|---|---|---|---|---|---|---|---|
| nPre | Level | nPost | nPre | Level | nPost | nPre | Level | nPost |
| 13 | ≥1 | 23 | 22 | ≥1 | 22 | 0 | =0 | 4 |
| 4 | ≥1.5 | 16 | 4 | ≥1.5 | 10 | 2 | ≥1 | 23 |
| 0 | ≥2 | 5 | 0 | ≥2 | 4 | 0 | ≥2 | 13 |
| 0 | ≥2.5 | 0 | 0 | ≥2.5 | 2 | 0 | =3 | 1 |

| Nature of an Experiment | | | Guiding Ideas and Questions | | | Results and Evaluation | | |
|---|---|---|---|---|---|---|---|---|
| nPre | Level | nPost | nPre | Level | nPost | nPre | Level | nPost |
| 4 | =0 | 4 | 11 | ≥1 | 25 | 18 | ≥1 | 25 |
| 23 | ≥1 | 23 | 1 | ≥1.5 | 15 | 10 | ≥1.5 | 19 |
| 3 | ≥2 | 14 | 0 | ≥2 | 7 | 5 | ≥2 | 11 |
| 1 | =3 | 4 | 0 | ≥2.5 | 2 | 2 | ≥2.5 | 8 |

**Table 4. Number of students scoring equal to or above coding levels overall and by section**

Level 2 answers focus on the development of a mechanistic understanding of the world. The point of science is to find out or discover how things work, where things come from, how things might behave in the future. Scientists have ideas, questions and predictions about how things work. Typical student answers often include Level 1-type statements, but in addition refer to more specific goals such as "to find out how something works," or "to find out how animals get oxygen." Scientists have ideas or questions about how things work or "how we got here," and predictions about "what will happen in a certain experiment."

Level 3 answers focus on the development of an explanatory understanding of the world. The point of science is to construct explanations for why things happen; scientists have ideas and questions about why things happen. In typical student answers, the goal of science is to find out "why something happens," "why the leaves change color," "why the dinosaurs became extinct," or "why the yeast bubbled in the tube."

The average pre-interview score on this section was 1.09 (Table 3). Scores ranged from .33 to 1.67. The median score was 1.0; the modal scores were 1.0 (n=8) and 1.33 (n=8). On the pre-interview then, students saw the purpose of science as discovering facts, making inventions, and developing cures.

There was a slight, significant increase in the mean scores on the post-interview (Table 3). As in the pre-interview, the average score hovered slightly above 1 (1.28), and the median and modal scores (n=11) were again 1.0. The range was extended -- from .33 to 3.0. Whereas no students scored 2 or higher on the pre-interview, 4 scored 2 or higher on the post-interview (Table 4, Fisher Exact Test, p<.06).

*(2) The Nature of a Hypothesis.* The question included in this section is: What is a hypothesis?

In Level 1, a hypothesis is an idea or guess. Typical answers were vague; a hypothesis is "an idea about something," or an "educated guess."

In Level 2, a hypothesis is also an idea or guess, but it is explicitly related to an experiment or phenomenon. The hypothesis is something that can be tested, or it is a prediction about an experiment or phenomenon. In typical Level 2 answers, a hypothesis is "an idea about something, but you can test it," or "an if-then statement about what you think might happen," or "an idea of what will happen in an experiment."

In Level 3, the hypothesis is not only related to an experiment, but aids in interpreting the results of the experiment, or is evaluated in terms of the results. The only Level 3 answer was given in the post-interview, and is quoted here at length, along with follow-up questions and answers.

> **Student**: [A hypothesis is] when you're trying to figure it out...like the guy [Linus Pauling] who was tipping the box on the screen...he thought it was a round thing, a cylinder. He was tipping the box and rolling it and he said it was a cylinder.
>
> **Interviewer**: Does he have just one hypothesis?
>
> **Student**: No, at the beginning he said it could be a cylinder with even sides or uneven sides, but when he tipped it, it made the same noise, the sides, so he knew it was a cylinder with even ends...his conclusion was that it was a cylinder with equal sides.
>
> **Interviewer**: Why did he have the hypothesis?
>
> **Student**: So you would know if what you would expect, but if you get the wrong answer, than you expected, you'd do it again, and if you got out the same answer, you'd know it was that, not the thing you thought it was gonna be.

Only two students were able to answer this question on the pre-interview; each gave a Level 1 answer. The remaining students did not know the word "hypothesis," and so received no score. All students answered on the post-interview, where the mean score was 1.37 (Table 3), with a range from 0 (n=4) to 3 (n=1). On the post-interview, the median score on the post-interview was 1; the modal score was 2 (n=12).

*(3) The Nature/Purpose of an Experiment.* The questions included in this section are: What is an experiment? Why do scientists do experiments? (Follow-up if answer is "To test ideas") How does the test tell the scientist something about the idea?

In Level 0 (presented here because the misconception it contains appeared so frequently), an experiment is described as a disembodied process. It is an activity that is not guided by an idea, a question, or an implicit assumption. Typical Level 0 answers include statements like, "an experiment is when you mix two chemicals together and see what happens," or "when you try something new."

In Level 1, there is no clear distinction between experiments and ideas. The motivation for doing an experiment, implied or explicitly, is to find something out about the thing being experimented with. In typical Level 1 answers, a scientist "tries something" to see if it "works" or "reacts," or to find something out. In an experiment, a scientist "takes samples to find out more about it." Scientists do experiments, "to check it out... find out about the thing they're experimenting on...[checking it out is] taking a good look at it...when you experiment something, you do something to it, like if you're gonna experiment of putting steel into water, you're putting steel into water, that's the experiment, finding out."

In Level 2, the distinction between the idea and activity is clear. The experiment is a test of a scientist's idea, or an operationalized exploration of an phenomenon. In typical Level 2 answers, scientists do experiments "to test to see if their idea is right" or "to test their hypothesis." An experiment is "trying to figure out your hypothesis...like yeast, you wanna figure out if it's alive. You cook it, then you put it in a [flask] and then you put the rubber stopper on and then if it gasses out it's probably alive and if it doesn't do anything, it isn't." (Note: although the student's interpretation of the possible results is backwards, it is clear that he or she understands that the experiment is motivated by a question.)

In Level 3, there is a distinction between the idea and the activity, and the experiment plays the same role as it does in Level 2. In addition, however, the relationship between the results of an experiment (especially unexpected ones) and the idea being tested is clearly articulated. Results aid in the evaluation or development of an idea, and ideas may change as a result of the work a scientist does. In a Level 3 answer, scientists do experiments "to test what they were thinking of...if it was the same thing they guessed it tells them they were right, and if it was different, it tells them they were wrong, and they need to change their idea."

On the pre-interview, the mean score was 1.0 (Table 3), with a range from 0 (n=4) to 3 (n=1). The median was1.0, as was the mode (n=20). On the pre-interview, then, students saw experiments as activities that support the goal of science -- finding things out and discovering facts.

The considerable increase in the mean score to 1.52 on the post-interview was significant (Table 3). Mean scores ranged from 0 (n=4) to 3 (n=4), and the median and modal (n=10) scores were now 2.0. By the post-interview, then, over half of the students saw experiments as tests of ideas, and some could articulate how unexpected results lead to revisions of ideas. This stands in marked contrast to the pre-interview, where only 3 students attained a score of 2 or higher (Table 4, $X^2$=8.90, $p<.005$).

(4) *Guiding Ideas and Questions*. The questions included in this section are: How do scientists do this [their] work? Where does a scientist get a hypothesis? How does a scientist decide which experiment to do?

In Level 0, there is no sense that the scientist is seeking information or has any other guiding purpose for the activities that he or she pursues, and no sense of the relationship between what the scientist does or thinks and anything other than the scientist's own whims and desires. A scientist does his/her work by "reading," "doing experiments," "doing research." A scientist does a certain experiment because he or she "feels like it."

In Level 1, the focus is on activities such as thinking, observing, or exploring, and the goal of these activities is to gather information. In typical Level 1 answers, scientists do their work by "looking it up in books," "putting things under microscopes to see how they behave,"and asking other scientists. They get a hypothesis "from their mind."

In Level 2, exploration is guided by a particular idea, question, object, or phenomenon. In a typical student answer, a scientist "walks through a forest and finds something new and tries to find out more about it." Another scientist "goes and looks for things. For example, if he goes to the moon, he finds a rock, he brings it to earth and experiments. [I mean] He puts, he has to know if it's alive or dead, if it needs energy."

In Level 3 either the guided exploration of Level 2 is elaborated to include operationalization or reflection on prior knowledge and experience, or there is an understanding of evaluation and development of ideas (as discussed in Section 3). In a typical Level 3 answer, a scientist "probably thinks up an idea, and then he builds an experiment out of the idea, and if he's right or wrong he keeps building up more questions to see, to find out even more stuff than he knows." Scientist's choice of experiment is based on "...an idea...and the stuff he already knows."

On the pre-interview, the mean score was .65, with a range from 0 (n=3) to 1.5 (n=1). The median and modal scores (n=14) were .5. Thus, on the pre-interview, students revealed the misconception a scientist's choice of hypotheses and experiments is mostly capricious.

The improvement on the post-interview was dramatic and significant (Table 3). The mean rose to 1.45, with a range from .33 (n=2) to 2.5 (n=2). The median score was now 1.33, and the modal score 1 (n=8). Fifteen students scored 1.5 or better, as opposed to only 1 on the pre-interview (Table 4, $X^2$=17.42, p<.001). While no students scored 2 or better on the pre-interview, seven did so on the post (Table 4, Yates' $X^2$ =5.91, p<.02). Thus, by the post-interview, most students understood that hypotheses and experiments are part of a focused effort to understand a particular question or phenomenon.

A similarly clear picture emerges when one considers the mean high scores, which improved significantly from near 1 on the pre-interview to over 2 on the post-interview (Table 4).

*(5) Results and Evaluation.* The questions included in this section are: Do scientists change their ideas? When or why? What happens when a scientist is testing an idea and gets a different result from the one s/he expected?

In Level 1, the scientist is trying to get a result; if the experiment doesn't come out "right," it is because something is not working and should be checked or changed. The "something" is not clearly specified as an idea. In typical Level 1 answers, the scientist "checks it to see what he did wrong and tries to fix the problem," or changes the idea/experiment a little by "adding stuff or taking stuff away."

In Level 2, the scientist is testing an idea; if the results of the experiment are unexpected, then, as in Level 1, something requires attention. In Level 2, however, the idea and experiment are clearly distinguished. In a typical Level 2 answer, "the person might think something went wrong in the way they did the experiment. so they go back to fix it," or, "they would change their idea."

In Level 3, there is an understanding that an idea is modified because of a conflict between the idea and experimental results or other evidence, and the modified idea takes these data into account. In a post-interview Level 3 answer, "he'd probably test it one more time and see if it came up with that again, and if he did, he'd probably have to change his hypothesis a little to fit in with the new data." The crucial part of this answer is the notion that the modified hypothesis must "fit" the experimental data.

The mean for the pre-interview was 1.06 (Table 3), with a range from 0 (n=2) to 3 (n=1). The median score was 1.0, as was the mode (n=9). Nine students scored less than 1, thirteen scored between 1 and 2, and five scored 2 or better (Table 4). While there was one student who demonstrated the understanding that a conflict between results ideas may lead to revisions in the ideas, the majority gave Level 1 responses in which even hypotheses and experiments are not clearly differentiated.

The improvement in the mean to 1.8 on the post-interview was dramatic and significant (Table 3). Scores for the post-interview ranged from .5 (n=2) to 3 (n=7). The median was 1.5, as was the mode (n=8). While ten students scored 1.5 or better on the pre-interview, nineteen did so on the post-interview (Table 4, $X^2$=6.04, p<.02). In contrast with the two students who scored 2.5 or better on the pre-interview, eight students scored 2.5 or better on the post-interview (Table 4, Fisher Exact Test, p<.04), demonstrating a clear understanding of the distinction between idea and experiment, and in some cases, of the relationship between idea and results.

*(6) Relationships.* The questions included in this section are: Do scientists do anything with their ideas? (Follow-up question if answer is "Test them":) How do scientists test their ideas, and what happens to the ideas once they've been tested? Is there a relationship between a scientist's ideas and the rest of the work a scientist does? What is the relationship? Is there a relationship between a scientist's hypotheses and the rest of the work a scientist does? What is the relationship?

The levels in this section are similar to those in the experiment section. We coded the two sections separately because they address the same issue (the relationship between ideas and experiments) from opposite points of view. We wanted to catch any answer differences that resulted from this difference in approach.

In Level 0, there is no relationship between a scientist's ideas and the rest of the work he or she does. Students at this level typically answer that there is no relationship, or else give a very incomplete rendering of it; scientists "report their ideas," or "write them down." It was sometimes difficult for us to gauge whether a student really had no understanding of the relationship, or whether he or she did not fully understand the questions.

In Level 1, there is no clear distinction between ideas and experiment. A scientist tries an idea to see if it works, "does it," or uses it as a guide or a blueprint. In typical Level 1 answers, scientists "make their ideas work," "see if they're accurate, if they can really do them," or "fulfill them by experimenting on them."

In Level 2, there is a clear distinction between the idea and the experiment. The idea is tested, to see if it's right, or the idea is used to predict the outcome of an experiment. In a typical Level 2 answer, "they test them [ideas]...in experiments, and see if they're right."

A basic Level 3 response goes beyond the Level 2 notion that ideas are tested in experiments to include the understanding that they are evaluated or developed in accordance with the results of these tests. A full Level 3 response would include, in addition to this, an understanding that the modification of an idea may entail reorganizing and re-interpreting the data on which the idea was originally based. None of the students gave such a full response, and students were given Level 3 credit for basic responses, such as the post-interview discussion quoted here:

> **Student**: Once they think up an idea, they experiment to find out more and then they probably come up with new questions to experiment with and they keep doing it until they find enough information on the one topic.
>
> **Interviewer**: How do scientists do experiments?
>
> **Student**: Like, with the yeast, if they wanna find out if it's alive or dead, they cook it at a certain amount for so long and then with different yeast they freeze it. Then they mix it with sugar and water, and if it bubbles, it probably means it's alive and if it doesn't do anything it means it's not.
>
> **Interviewer**: So what happens to the ideas once they've been tested?
>
> **Student**: Most of the time they change them.
>
> **Interviewer**: Why?
>
> **Student**: Because their first idea might be wrong.
>
> **Interviewer**: How would they know it was wrong?
>
> **Student**: Like, maybe they would think of an idea, like that your name is something else, and they do much more research and they find out what your real name is.

The mean score on the pre-interview was .91 (Table 3); the range from 0 (n=7) to 3 (n=1). The median and modal (n=16) scores were 1.0. Only three students scored 2 or better (Table 4). Thus, the overwhelming majority of the students saw ideas, at best, as blueprints for action, or interchangeable with the things they are about.

The improvement in the mean score to 1.69 on the post-interview was dramatic and significant (Table 3). The median score was now 1.5, and the modal score 2.5 (n=8). Fourteen of the subjects scored 2 or better on the post-interview (Table 4, $X^2$=22.74, p<.001).

Again, mean high scores increased from around 1 to over 2, which is also significant (Table 3).

*Discussion*

*Seventh Graders' Initial Understanding.* The overall mean on the pre-interview was 1.0. Only four students had overall mean scores over 1.5, and there were only eleven scores of 2 or more scattered through the six sections of the interview. According to our instrument, then, the seventh graders in the study may be characterized as having a Level 1 understanding of the nature of science and scientific inquiry prior to instruction.

Perhaps the critical feature of Level 1 is the absence of an appreciation that ideas are distinct, constructed and manipulable entities. There is no understanding that a scientist's ideas motivate the scientist's other, perhaps more tangible work, such as gathering data and doing experiments, or that the ideas, in turn, are affected by this work. Instead, ideas are confused with experiments, or with whatever else they are about (an invention, cure, and so on), and there is no acknowledgement of the theoretical motivations behind scientists' experiments and other activities.

More generally, in Level 1 understanding, nature is there for the knowing. Accordingly, one of the goals of science is to discover facts and answers that exist, almost as objects, "out there." In typical Level 1 fashion, there is no understanding that "facts" and "answers" are actually constructed ideas *about* objects "out there." Other goals include inventing new things and finding cures for diseases. Here, too, the idea is equated with the thing, or else a simple plan of action ("They have an idea for a rocketship, so they do it"). Scientists work towards their goals by observing things and looking for discoveries, or by trying things out to see if they work, but are never explicitly guided by an idea.

*The Effect of the Curriculum.* Overall, the per-item mean increased from 1.0 to 1.55 on the post-interview. This increase was statistically significant (p<.001, Wilcoxon signed ranks test, 1-tailed). Every child improved, and improvement averaged half a level. Now 16 students achieved overall scores of 1.5 or better, and eight scored 1.9 or better on the post-interview, a score nobody achieved on the pre-interview. There were 41 scores of 2 or better on the various sections of the interview. These results stand in marked contrast to those on the written test. The interview is clearly a more sensitive instrument for assessing the effects of the Nature of Science Unit.

The greatest score increases occurred in the ections on Guiding Ideas and Questions, Results and Evaluation, and Relationships. Viewed in terms of the Nature of Science Unit, these results make sense. While the Unit did incorporate lessons that focused specifically on hypotheses and experiments, its emphasis was on relationships among these elements in the development of scientific ideas, and the highest scoring sections of the interview all made reference to these relationships.

The main points of the unit's introductory lesson were that the basis for scientific inquiry is mental work, and that experiments are tests of ideas. These points were put into practice in the black box lessons and the yeast lessons, where the students conducted experiments or tests only *after* they had explicitly formulated a hypothesis about a certain problem or phenomenon. The Guiding Ideas and Questions section of the interview, in which the mean score increased from .65 to 1.45, emphasized this relationship between ideas and experiments.

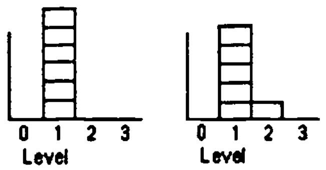
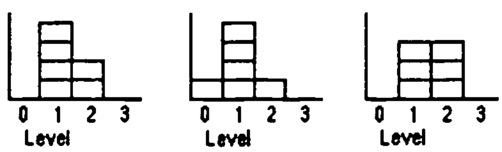
In both the black box and the yeast lessons, students practiced evaluating their ideas by looking at the experimental data that they had generated. The yeast lessons made the additional point that new ideas and questions may arise from the results of an experiment. The Results and Evaluation section of the interview, in which the mean score increased from slightly above Level 1 almost to Level 2, emphasized the relationship between an idea and the results of an experiment.

The Relationships section of the interview emphasized the role of ideas in motivating and guiding experiments, and the role of data and results in evaluating ideas. In these questions, students were essentially asked to integrate the points that had been made throughout the Unit. The mean score increased from .91 to 1.69.

*Consistency of Scores Across Sections.* In order to address the question of how consistent students were in their scores across the six sections, we noted the highest score a subject received on each section, and analyzed how tightly clustered these scores were. We found the clustering to be moderate; see Table 5 (next page).

In the pre-interview, over half of the students received the same score for all six sections (in every case, this was Level 1), with at most 1 section receiving the adjacent score (Level 0 or Level 2). Even on the post-interview, half of the subjects located their high scores on one main level, or on two adjacent levels. On the post-interview, these almost always included Level 2 or Level 3 responses. Consistency was greater on the pre-interview than on the post-interview because of the preponderance of Level 0 and Level 1 responses on the pre-interview, and because the curricular intervention affected the students' understanding of the material in some sections more than in others.

*Are the Levels "Stages?"* Are the levels stages, as in Kohlberg's stages of moral understanding? No claims are here being made that the interview places students into stages. Before this possibility could even be systematically explored, more careful articulation of the differences between levels (especially between Level 2 and Level 3) would be needed. The scoring procedures would require refinement, so that interscorer reliability could be improved over our current 75 percent level. So, while the present effort indicates that it might be possible to identify stages in the understanding of the nature of science and scientific inquiry, it is only a first step towards doing so.

| Clustering of sections by level score | n pre | n post |
|---|---|---|
| Single dominant level  | 14 | 2 |
| Tightly clustered distribution  | 3 | 11 |
| Other | 10 | 14 |

▭ = 1 interview section

**Table 5. Clustering Patterns of High Scores on the Six Sections of the Study 2 Clinical Interview**

## Observations and Teacher and Student Reactions to the Revised Unit

In general, both the teacher and the students enjoyed the lessons and felt they were appropriately educational. Table 6 shows student reactions to this effect.

Below, we report the observations and reactions of the teacher, students, and observing group members, for all of the lessons.

### Introductory Lesson

The students were very curious about the burnt-looking "martian object" under a plexiglass dome in front of them. Throughout the class, they speculated about what it might be: a rock, an "orbitee," a burnt brownie, and so on. They reflected on the fact that their ideas about what it means to be "alive," like scientists' ideas, came from their own experience and resided in their own minds. When the teacher asked how they could find out whether the object possessed any of the attributes of life that they had listed, some students responded, "Do an experiment!" Although they seemed to understand the logic of testing ideas, they came up with no examples of well-designed experiments, and were curious to find out how they could test their ideas.

| | Very Good | Good | Fair | Poor | Unsatis-factory |
|---|---|---|---|---|---|
| Animal Mimicry | 10% | 74% | 13% | 3% | ---- |
| Black Box Problem | 6% | 27% | 53% | 15% | ---- |
| Yeast | 73% | 22% | 3% | ---- | ---- |
| Entire Unit | 23% | 58% | 19% | ---- | ---- |

**Table 6. Student Reactions Regarding Educational Appropriateness and Enjoyment**

Videodisc Lessons

*Animal Mimicry Lessons*

Whereas the main point of the introductory lesson was made quite clear to the students, the main points of the animal mimicry lessons, unfortunately, were not as clearly articulated. The students were engaged by the videodisc, but did not understand the homework or the reasons for subsequent tasks. Some of the students did not understand the concept of "category," and a lack of class time did not allow for thorough student discussion. When asked where the categories came from some students answered, "from books and computers," even after the teacher had pointed out that the categories came from the students own minds and the observations they made. In addition, students did not see beyond the categorizing task to the usefulness of the categories in building an explanation for animal disguise. It was decided that if the lessons were to be used again, they should include another day of general discussion about categories; what they are, and why they are useful.

*The Black Box Problem*

As was predicted, when students were asked for suggestions as to how to go about finding the shape of the object inside the enclosed box, they suggested randomly shaking it or moving it in one direction or another. After this exercise, students watched the video-clips of Dr. Pauling methodically testing for various aspects of the objects shape. They enjoyed guessing which hypothesized shapes were supported and not supported by the results of Dr. Pauling's tests.

For homework, to the surprise of the teacher, most students diligently tested for the shape of an object. On the second day, however, the students had difficulty understanding why they had to record what they were doing, one step at a time. In answer to studer' questions, the teacher usually replied, "If you just shook the box up, how would you know what was inside of it? But if you test one aspect of it at a time, you can find something out about it which will help you know more about your problem." All students seemed to accept t' explanation.

Most students were disappointed to find out that they were not actually going to find out what the objects were in each of their boxes. There was not enough time at the end of the lessons to discuss why the absence of a final, "right" answer was acceptable, and similar to the lack of certainty in the solutions to "real" scientific problems. Again, both the teacher and research group felt that an extra day of instruction would have been beneficial. Even so, many students referred to the instructional points made during these two days in their post-interviews.

Yeast Lessons

In follow-up questions to the post-interview, the majority of students stated that they enjoyed the yeast lessons the most. In fact, throughout these lessons, all of the students showed a great amount of exuberance for "doing experiments."

The content of the lessons provoked student curiousity. The issues raised in the yeast problem gave the teacher a number of opportunities to draw on students' pre-instructional knowledge about bread and the function of yeast. One of the differences between these revised lessons and the earlier version was that more time was allocated for student questions and detailed discussions. Given this opportunity, students tended to be highly engaged, animated, and often gave freely of their ideas.

It seems as though the homework questions that related the experimental evidence to real world problems worked well. Students were able to use their common sense/everyday knowledge about bread making along with evidence from the experiments to form theories about the reasons for the effect of temperature on bread dough.

Observations of the yeast wrap-up lesson confirmed our thoughts that these lessons successfully communicated the points we wanted to make about science as a process of inquiry in which scientists seek increasingly deeper explanations of natural phenomena. At the beginning of the day, the students were able to offer simple explanations about the yeast phenomena. Examples of their statements are: "If y-- freeze it it dies; if you bake it it won't"; "Yeast is alive but t' can't live in hot places"; and, "Frozen yeast makes bubbles when mixed because it's preserved. Baked yeast doesn't because the nutrients or something are taken out of it." Through a subsequent discussion of what they know abou- similar phenomena such as hibernation, students began to shape an explanation of the mechanism -- that when frozen, yeast may become dormant. This lead them to a discussion of the verification experiment: if we warm the mixture that incuded frozen yeast, will it become active again? Some students were able to ge.. ate predictions about what would happen, and able to adequately interpre. and explain the results of their experiments.

Students were able to generate a number of plausible explanations for a single question that they then agreed could be tested: Why doesn't the dough keep rising? Three reasons were concluded: (1) the chemical reaction of yeast, sugar and water wears out; (2) heat kills the yeast; and, (3) the yeast runs out of sugar or energy ("If you pour in a bucket of sugar and put it in the oven it would keep rising forever"). Why doesn't frozen dough rise?: "They're dormant"; "Yeast that freezes dies"; and, "The liquid in the cells are frozen so it can't take in the sugar." Why do we put dough in a warm place to rise?: "(At warm temperatures) it starts to come out of it's dormancy and if you heat it too much it will die" and "it's in hibernation; warmth brings it out."

Wrap-up Lesson

Students became very active in discussing the various lessons of the unit and how they approached each problem presented to them. This was especially true when the teacher brought up the yeast experiments. Students talked actively about how the questions had changed and become deeper during the course of the week's activities. The teacher raised some intriguing questions: "How do they make yeast?" and "Why is there an expiration date on the package?" In reaction to her questions, many students offered their own: "What happens when you put yeast in a microwave oven?" and "What happens when you use Sweet-and-Low?"

**Conclusions to Study 2**

Overall, the teacher and the students enjoyed and profited from the revised Nature of Science Unit. The yeast lessons were successful and much more manageable, both logistically and conceptually, than they had been in Study 1. Although the students had difficulty understanding some of the task requirements of the videodisc lessons, they did in fact often refer to the activities from the black box lessons when discussing various points about scientific inquiry in the post-interviews. We take this as a sign that in spite of the difficulties, these lessons also succeeded in communicating some of the unit's points. However, we cannot determine the degree to which use of the "Seeing the Unseen" videodisc contributed to the communication of these points. Alternative lessons without the use of the "Seeing the Unseen" videodisc might be successful in communicating these same points, but only further research can determine what unique role, if any, interactive videodiscs might play in the teaching of such lessons.

The revised clinical interview elicited students' understanding of the nature and process of scientific inquiry and measured the effects of the unit more effectively than did the Study 1 version of the interview. It is interesting to note that while every section on the interview showed statistically significant improvement pre- to post-, this was not true for the written test. For the two sections of the interview where students showed the greatest improvement in mean score -- Guiding Ideas and Questions and Results and Evaluation -- there was no such improvement for the corresponding sections in the written test (in fact, there was a small, non-significant drop on the Results and Evaluation section). There are several possible explanations for this discrepancy. One is that is that the interview and the written test assessed different areas of the students' understanding. Another is that the students' understanding was drawn out more fully in conversation than it could be on a multiple choice test, thus indicating that the interview is a more sensitive instrument than the written test. A final possible explanation for the differences between interview and written test scores on certain sections is that the interview actually lead the students to a clearer understanding of the material than they reached without the probing of the interviewer.

## OVERALL CONCLUSIONS

Our intentions in these two studies were to assess what seventh graders understand about the nature of science and inquiry, and to explore our ideas about what should and can be taught about these issues at this level, including giving our curricular lessons two field trials. Here we will discuss several issues -- the importance of teaching students about the nature and process of scientific inquiry, and what we have learned about what is and is not effective in doing this, including the role of technology.

Of what use is it to junior high school students, or senior high school students for that matter, to learn about the nature of scientific inquiry? A large part of the justification follows from an already accepted goal of science education -- to teach about science as an enterprise. It is important for citizens to be empowered to reason critically about all knowledge, including purported scientific knowledge. Having gone through the process of theory building themselves, we hope that students will be in a better position to understand that the body of scientific knowledge is constructed and changing, rather than the objective and final "truth." They will be able to evaluate the reasoning that supports this knowledge; to ask where a piece of knowledge comes from, and what the evidence for it is.

In order to do this, students must have an appreciation that scientists' ideas are constructed and manipulable entities, distinct from experiments, data, and the object or phenomenon being studied, and they must recognize the particular role of experiments and data in theory construction. As evidenced by the results of the Study 2 clinical interview, what characterizes seventh graders' initial understanding of the nature of science and inquiry is precisely the absence of such a perspective. What have we learned about what makes curriculum successful at moving students beyond their initial understanding?

Such a curriculum should contain problems that allow students to construct and manipulate their own ideas about a phenomenon. We think that a good problem is one about which students have some intuitive ideas that can be worked with and developed in a sustained way. The problem itself should be motivated by an observation that provokes a question such as Why is something the way it is, or Why is something not as we expect it to be? In investigating the problem, it is equally important that *each step of the process be motivated.* Rather than simply being presented by the teacher or the textbook, hypothesized explanations should be motivated by a discussion of what students already know about this or related phenomena. Likewise, the choice of variables and the design of experiments should be motivated by a consideration of students' ideas about what is relevant to the problem and to the hypothesis being tested. Possible outcomes of an experiment, and what each would show about the idea being tested, should be discussed before the experiment is actually performed. This motivates students to objectively evaluate their hypotheses, and to reformulate them if confronted with contradictory evidence.

What have we learned about how to structure this sort of curriculum? Throughout the curriculum, it is crucial that students are asked to explicitly reflect upon why they are doing what they are doing. This necessitates setting aside sufficient time in all lessons for class discussion, particularly before and after experiments. We have also found that experiments do not need to be hands-on in order to effectively communicate points about the process of inquiry. A demonstration can be quite valuable, as long as the logic of the experiment is made clear ahead of time through careful discussion of the motivation for the experiment and the meanings of the possible results. For example, in the revised yeast lessons, the "bromthymol blue"experiment was done as a classroom demonstration. Bromthymol blue is a solution that changes from blue to yellow in the presence of $CO_2$. While the specific purpose of the experiment was to test for the presence of $CO_2$ in the gas given off by yeast reaction, the more general purpose was to add to the growing body of evidence that ultimately supported the hypothesis that yeast is alive. The teacher made both of these purposes quite clear at every step along the way, and discussed what each of the possible outcomes of the experiment would mean for the competing hypotheses <u>before</u> she carried out the experiment. In class discussions, the students were able to articulate the logic behind the experiment and to interpret the results of the experiment without having manipulated any of the materials themselves.

In the process of further revising the Nature of Science Unit, the two software pieces from Study 1 were dropped, as were the animal mimicry lessons from Study 2. The above discussion makes it clear why these were the least successful of our curricular materials. In none of these lessons did students encounter a phenomenon that provoked a question and motivated a sustained process of building new scientific understanding. Instead, they worked in piecemeal fashion on seemingly arbitrary problems. The element of *motivation*, which was essential to the successful sections of our unit, was missing.

Does this mean that there is no role for technology in imparting points about the nature of scientific inquiry? Not at all. Whenever technology is well suited to actually building scientific understanding, it is well suited to making points about the enterprise. So for example, White and Horowitz (1987) developed a microworld that allows students to discover Newton's laws, and included curricular materials about the nature of laws, the criteria for deciding whether something is a good law, and so on. Similarly, two ETC projects (Weight and Density; Heat and Temperature) use the representational capabilities of computers to provide interactive models of natural phenomena. Points about the nature of models, including discussions of what makes a good model, the revisability of models, and the existence of many models that are useful in helping us understand any given phenomenon, have been incorporated into these curricula.

Science learning, at any age, involves knowledge restructuring and conceptual change; these do not come easily. An additional aim of the Nature of Science Unit is based on our contention (not tested in these studies) that if students understand something of the process of developing scientific knowledge and the constructed nature of this knowledge, they will be in a position to understand the evidence that will lead them to restructure their own knowledge. We hope to address this question in the future.

**References**

Carey, S., Honda, M., Jay, E., O'Neil, W. and Unger, C. (1986). What Junior High School Students Do, Can, and Should Know About the Nature of Science. Educational Technology Center Progress Report #86-11. Cambridge, MA: Harvard Graduate School of Education.

Chomsky, C., Honda, M., O'Neil, W., and Unger, C. (1985). Doing Science: Constructing Scientific Theories as an Introduction to Scientific Method. Educational Technology Center Technical Report #85-23. Cambridge, MA: Harvard Graduate School of Education.

Cronin, L.L. and Padilla, M.J. (1986). The development of a middle grades integrated science process skills test. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, March 1986.

Fraser, B.J. (1979). Test of Enquiry Skills: Handbook. Victoria: The Australian Council for Educational Research Limited.

Klopfer, L.E. and Carrier, E.O. (1970). TOUS: Test on Understanding Science (Form Jw). Pittsburgh, PA: Learning Research and Development Center, University of Pittburgh.

Storey, K. and Mellin, C. (1987). A Prototype Science Interactive Videodisc: Research on In-School Use. Educational Technology Center Technical Report #87-4. Cambridge, MA: Harvard Graduate School of Education.

White, B.Y. and Horwitz, P. (1987). ThinkerTools: Enabling Children to Understand Physical Laws. BBN Laboratories Inc. Report #6470. Cambridge, MA: BBN Laboratories.

Appendix 1

Study 1 Written Test

Name: ______________________________

Group and Number: ______________________________

Directions:
This is a test to find out what you know about science and the scientific method. Read each problem carefully. Choose what you think is the BEST answer, and circle the answer. Try hard to answer all of the questions.

1. When a scientist constructs a new scientific theory, we know that she or he has

   a. revealed one of the laws of nature.

   b. helped to move mankind closer to absolute truth.

   c. discovered new ways to conduct experiments.

   d. developed new ideas and understanding.

2. After scientists think they have found the solution to a problem in scientific research, they usually

   a. stop doing science because they have achieved their goal.

   b. see that their work has opened up new problems.

   c. cannot easily find a new problem to work on.

   d. feel that the problem is solved once and for all.

3. When much new evidence is discovered which does not fit a scientific theory, scientists usually will ask themselves:

   a. Shall we throw out the theory, since the new evidence doesn't fit it?

   b. Can we change the evidence a little so that it will fit the theory?

   c. Shall we keep the theory as it is, since this new evidence doesn't help it?

   d. Can we change the theory a little so that this new evidence will fit it?

4. Ralph said, "Scientists do experiments to ask questions of nature." By this, Ralph means that experiments are used in science to

   a. prove that nature works in a regular manner.

   b. learn by trying things, making mistakes, and then trying again.

   c. test out the ideas of scientists.

   d. ask about the mystery of creation.

In 5, circle whether each statement is TRUE or FALSE.

5. Measurement is important in science because

   a. measurement improves the precision of our observations. TRUE FALSE

   b. many predictions are about numerical relations between variables--for example, the more you exercise the faster your heart beats. TRUE FALSE

   c. without measurement scientists would have nothing to do. TRUE FALSE

6. A scientific <u>theory</u> should

   a. give final answers to scientific questions.

   b. supply directions for making useful things.

   c. tie together and explain many related natural events.

   d. describe the world as most scientists see it.

7. In forming a hypothesis, a scientist tries to develop an explanation that best fits

   a. as much of the data as possible

   b. more than half the data

   c. only the most recent data collected

   d. only the data collected in his or her laboratory

Questions 8 - 11 refer to the following situation.

A fire destroyed a large forest, and the soil is being washed away by rain. So, the forest rangers want to test how different types of grasses affect soil erosion. They choose ten plots of ground that are the same size. These plots receive the same amount of sun. They also have the same kind of soil. The rangers plant each plot with a different type of grass. Measurements of soil erosion are made every week for the entire summer.

8. What is the independent variable?

   a. the size of the plots

   b. the types of grasses

   c. the amount of soil erosion

   d. the type of soil in the plots

9. What is the dependent variable?

   a. the size of the plots

   b. the types of grasses

   c. the amount of soil erosion

   d. the type of soil in the plots

10. What hypothesis is being tested in this study?

    a. Some types of grass reduce soil erosion more than others.

    b. Soil erosion is affected by the slope of the land.

    c. Burned over areas have greater erosion than forested areas.

    d. Planting grass will reduce the amount of soil erosion.

11. Which of the following variables is NOT controlled in this study?

    a. the size of the plots

    b. the type of soil in the plots

    c. the amount of soil erosion

    d. the amount of sun the plots received

12. The scientists of today can work on more complex problems than the scientists of the past because they

    a. work much harder than earlier scientists.

    b. have more imagination than earlier scientists.

    c. build on work of earlier scientists.

    d. are more intelligent than earlier scientists.

13. Most scientists consider scientific theories to be

    a. changeless truths.

    b. explanations that may be revised.

    c. final explanations.

    d. descriptions of the world as it really is.

14. Scientists make progress by

    a. describing the work of other scientists.

    b. repeating experiments over and over.

    c. forming hypotheses and testing them.

    d. rejecting data from previous experiments.

15. All of the considerations below contribute to whether a problem is a good scientific problem. Which one is the LEAST important?

    a. The solution to the problem would help choose between two competing theories.

    b. The answer can be easily and quickly discovered.

    c. The solution to the problem would satisfy the individual scientist's curiosity.

    d. The problem is clearly stated and the methods for solving it are understood.

Questions 16 - 19 refer to the following situation.

John wanted to find out which laundry soap was best for removing grass stains. He tested four soaps. He mixed 1 tablespoon of each soap with 1 quart of warm water at 20 degrees Celcius. Each soapy mixture was then used to scrub a piece of a grass-stained cloth for 1 minute. Then the amount of stain left on the cloth was measured.

16. Of the following variables, which one was NOT controlled in this experiment?

    a. water temperature

    b. amount of stain left on the cloth

    c. amount of water

    d. amount of time scrubbing

    e. amount of laundry soap

17. Why did John need to control variables?

    a. He wanted the conditions of each trial to be the same so that he could see if the variable he was testing made a difference.

    b. By controlling variables, he would get exactly the same outcome for all four trials.

    c. He needed to control the independent variable because that was the variable he was trying to test.

    d. He needed to control things so that he would not make a mess.

18. Why was it important that John measured amounts, time, and temperature in his experiment?

    a. He wanted to find out if using more laundry soap would make the stain come out faster.

    b. If he had not made measurements, he could not be sure that the controlled variables were the same for each trial.

    c. He needed to practice measuring variables accurately.

    d. It did not matter if he measured the other variables, but one must always measure the dependent variable.

19. In the previous experiment, John mixed the soap and warm water in plastic containers. Two containers were made of blue plastic and two were made of white plastic. Failing to control for the colorof the containers was a serious scientific mistake. TRUE or FALSE?

    a. TRUE. He should have controlled for color.

    b. FALSE. He did not have to control for color.

20. Which one of the following does NOT result from the scientist's creative thinking?

    a. Theories, like modern chemistry.

    b. Concepts, like atom or density .

    c. Natural events, like the fact that solid objects fall to the ground when dropped.

    d. Inventions, like the microscope or the telescope.

21. Scientists predicted that an experiment would come out a certain way. They observed a different result when they did the experiment. What would they most likely say to themselves?

    a. "We shouldn't have made a prediction before trying this out."

    b. "Something went wrong in the experiment, because we didn't observe the result we predicted."

    c. "We should have better equipment for this experiment to get the right result."

    d. "Something is wrong either with our observation, our experiment, or our prediction."

22. Which one of the following statements about scientific theories is FALSE?

    a. A scientific theory gives a description of the world as it really is.

    b. A scientist may reject a scientific theory if she or he has doubts about it.

    c. When there is some evidence against a scientific theory, scientists may still accept it.

    d. A scientific theory gives an explanation o[illegible]ated things in nature.

23. Scientists have many goals in doing scientific work, but their principle aim is to

    a. search out errors in what has already been discovered about the physical universe.

    b. explain things in nature in terms of principles and theories.

    c. discover, collect, and classify facts about living and non-living things.

    d. provide the people of the world with the means for leading better lives.

Questions 24 and 25 refer to the following situation.

Konrad was very interested in Mallard ducklings (baby ducks). In particular, he liked to watch the way ducklings follow the mother duck around. To explain this, he made two competing ideas or hypotheses:

> Hypothesis 1: The <u>quacking</u> of the mother duck is what attracts Mallard ducklings to her.
>
> Hypothesis 2: The <u>movement</u> of the mother duck is what attracts Mallard ducklings to her.

To learn which, if either, hypothesis was right, Konrad made the observations listed below. Some of these observations support Hypothesis 1, some support Hypothesis 2 and some support both ideas. Some observations contradict (argue against) one or both of the hypotheses. And, some observations tell us nothing about either or both of the hypotheses.

24. For each of the following observations, fill in the blanks. Write down whether the observation SUPPORTS, CONTRADICTS or TELLS NOTHING about Hypothesis 1 and Hypothesis 2.

| <u>Observations</u> | <u>Hypothesis 1</u> (quacking) | <u>Hypothesis 2</u> (movement) |
|---|---|---|
| a. A white farmyard duck quacked like a Mallard duck and was followed by the ducklings when it moved. | ________ | ________ |
| b. The ducklings did NOT go near a silent toy Mallard duck which was not moving. | ________ | ________ |
| c. A duck, which could NOT quack, hatched some Mallard eggs and the ducklings followed her when she moved. | ________ | ________ |
| d. Konrad quacked like a Mallard duck and the ducklings followed him when he moved. | ________ | ________ |
| e. Konrad crowed like a rooster and the ducklings followed him when he moved. | ________ | ________ |
| f. A dog barked at the ducklings and they ran away. | ________ | ________ |

25. Keeping in mind your answers to the previous question, which of Konrad's hypotheses is better supported?

    a. Hypothesis 1: The quacking of the mother duck is what attracts the Mallard ducklings to her.

    b. Hypothesis 2: The movement of the mother duck is what attracts the Mallard ducklings to her.

    c. Neither hypothesis is better supported than the other.

26. Alice grows violets. She has six red and six white violets. She heard that violets produce more flowers when they receive morning sunlight. She made this hypothesis:

    Hypothesis: When violets receive morning sunlight rather than afternoon sunlight, they will produce more flowers.

    Which of the following is the BEST plan to test this hypothesis?

    a. Set all of her violets in the morning sun. Count the number of flowers produced by each. Do this for a period of four months. Then find the average number of flowers produced by each kind of plant.

    b. Set three white violets in the morning sun. Set the other three white violets in the afternoon sun. Do not study the red ones at all.

    c. Set all of her plants in the morning sun for four months. Count the number of flowers produced during this time. Then set all of the plants in the afternoon sun for four months. Count the number of flowers produced during this time.

    d. Set three red and three white violets in the morning sun. Place the other three red and three white violets in the afternoon sun. Count the number of flowers produced by each plant for four months.

Appendix 2

Study 1 Clinical Interview

## Interview Questions

1a. What do you think science is all about? (or What do you think is involved in science?)

1b. Why does a scientist do those things?

1c. What is the goal of science? (or What's the point of doing science?)

1d. Why do people so science?

---

2a. Have you ever heard of the word "theory"?

2b. What are theories all about?

2c. What does a theory do?

---

3a. How do you get a theory?

---

4. If you have a theory, what do you do with it?

---

5. Do you think scientists change their theories?
   If "yes": Why?
   If "no": Why not?
   If "some do and some don't": Why do the ones who do change their theories, change their theories? Any why do the ones who do not change their theories, not change their theories?
   If "it depends": It depends on what? ... Can you give me an example?

---

3b. How does a scientist get his or her ideas (or theories)?

3c. Where does a scientist get his or her ideas (or theories)?

---

6. How do scientists come to understand things they don't already know?

---

7a. Can you tell me what an experiment is?

7b. Why do scientists do experiment? [Prompt student to state the purpose.]
If "to prove": What do you mean by "prove"?

---

8. How does the scientist choose what experiment to do?

---

9a. Have you heard of the word hypothesis?

9b. Why have a hypothesis?

---

10. [Relate this question to any previous discussion about process or experimentation, testing things/ideas/theories/hypotheses.]
What happens if you get a different result than the one you expected?

Appendix 3

Quantitative Analysis of Study 1 Pre/Posttest Clinical Interview

The mean score per-item on the Study 1 clinical interview rose from .73 to 1.1 in the experimental group. This increase was significant ($p<.01$, Wilcoxon signed ranks test, 1-tailed). The increase from 1.01 to 1.39 in the control group was not significant. Scores rose from 1.07 to 1.55 in the above-average group (not significant), and from .69 to .98 in the below-average group ($p<.02$, Wilcoxon signed ranks test, 1-tailed). These results are reported in Table 2. It is worth noting that the pattern of significance is determined by sample size.

| | Pre | Post | Increase | |
|---|---|---|---|---|
| Overall | .85 | 1.22 | .37 | n=14 |
| Experimental | .73 | 1.10 | .37 | n=8 |
| Control | 1.01 | 1.39 | .38 | n=6 |
| Above Average | 1.07 | 1.55 | .48 | n=6 |
| Below Average | .69 | .98 | .29 | n=8 |

**Mean Score Per Item on the Study 1 Clinical Interview**

The overall increase in mean score per-item was significant for only two sections: The Role of Guiding Ideas and Questions, and The Nature/Purpose of Experiments. The mean score per-item for The Role of Guiding Ideas and Questions rose from .47 on the pre-interview to .96 on the post-interview ($p<.05$, Wilcoxon signed ranks test, 1-tailed). The mean score per-item for The Nature/Purpose of Experiments rose from .96 on the pre-interview to 1.43 on the post-interview ($p<.01$, Wilcoxon signed ranks test, 1-tailed).

The increase in mean score per-item in The Role of Guiding Ideas and Questions was not significant in either group alone. The increase in the mean score per-item in The Nature/Purpose of Experiments was significant in the experimental group ($p<.05$, Wilcoxon signed ranks test, 1-tailed), but not in the control group. The score in the experimental group rose from .75 to 1.5, and in the control group, from 1 to 1.33. It is not surprising that the increase was significant in the experimental rather than the control group, since the question addressed the aspects of an experiment (such as the relationship between experiments and ideas) that were emphasized in the experimental unit rather than those (such as the identification of variables) emphasized in the control unit.

Appendix 4

Study 2 Written Test

Name: ______________________________ Period: _____
Date: _______

Directions:
This is a test to find out what you know about doing science. Read each problem carefully. Choose what you think is the ONE BEST answer, and circle the answer. Try hard to answer all of the questions.

1. When scientists construct a scientific theory about an event in nature, we know that they have

   a. obtained the truth about the event being studied.

   b. discovered one of the laws of nature.

   c. uncovered the facts about the event simply from observing it.

   d. developed the best explanation of the event that they can have.

2. After scientists think they have found the solution to a problem by thinking it through and doing an experiment, they almost always

   a. use the solution to make an invention.

   b. see that the results of the experiment have led to new questions.

   c. have difficulty finding another problem to work on.

   d. feel that the problem is solved once and for all.

3. When new information is discovered that conflicts with a theory, scientists will usually

   a. throw out the theory, since the new evidence does not fit in.

   b. change the evidence a little so that it will fit the existing theory.

   c. keep the theory as it is, since this new evidence does not support it.

   d. change the theory a little so that this new evidence is explained by it.

Questions 4 and 5 refer to the following situation.

A fire destroyed a large forest, and the soil is being washed away by rain. So, the forest rangers want to test how different types of grasses affect soil erosion. They choose ten plots of ground that are the same size. These plots receive the same amount of sun. They also have the same kind of soil. The rangers plant each plot with a different type of grass. Measurements of soil erosion are made every week for the entire summer.

4. What idea is probably being tested in this study?

    a. Some types of grass reduce soil erosion more than others.

    b. Soil erosion is effected by the slope of the land.

    c. Burned over areas have greater erosion than forested areas.

    d. Planting grass will reduce the amount of soil erosion.

5. Which of the following variables is <u>NOT</u> controlled in this study?

    a. the size of the plots

    b. the type of soil in the plots

    c. the amount of soil erosion

    d. the amount of sun the plots received.

In the following questions, circle whether you think each statement is TRUE or FALSE.

| | | |
|---|---|---|
| 6. TRUE | FALSE | Scientists usually do not have any ideas about how something works before they start experimenting on it. |
| 7. TRUE | FALSE | For the entire time they spend investigating something, scientists are always working on the exact same question that they started out with. |
| 8. TRUE | FALSE | Scientists sometimes need to create a laboratory version of an event in nature so that they can observe it. |
| 9. TRUE | FALSE | Scientists can only investigate an object if they can see or touch it. |
| 10. TRUE | FALSE | As they investigate things in nature, sometimes scientists have to change basic ideas that they had once believed to be true. |

11. Scientists decide what variables to test in an experiment

    a. by using the variables that are given in the problem.

    b. by trying every variable that can be tested.

    c. by considering all the information available and their ideas about what is happening.

    d. by choosing any variable since it doesn't matter which one they start with.

Questions 12 - 14 refer to the following situation.

John wanted to find out which laundry soap was best for removing grass stains. He tested four soaps. He mixed 1 tablespoon of each soap with 1 quart of warm water at 20 degrees Celcius. Each soapy mixture was then used to scrub a piece of grass-stained cloth for 1 minute. Then the amount of stain left on the cloth was measured.

12. Of the following variables, which one was <u>NOT</u> controlled in this experiment?

    a. amount of stain left on the cloth
    b. amount of water
    c. amount of time scrubbing
    d. amount of laundry soap

13. Why did John need to control variables?

    a. He wanted the conditions of each test to be the same so that he could see if the variables he was testing made a difference.

    b. By controlling variables, he would get exactly the same outcome for all four tests.

    c. He needed to control only the variables that he was trying to test.

    d. He needed to control things so that he would not make a mess.

14. Why was it important that John measured amounts, time, and temperature in his experiment?

    a. He wanted to find out if using more laundry soap would make the stain come out faster.

    b. If he had not made measurements, he could not be sure that the controlled variables were the same for each test.

    c. He needed to practice measuring variables accurately.

    d. It did not matter if he measured the other variables, but one must always measure the outcome variable.

15. Scientists predicted that an experiment would come out a certain way. When they did the experiment, they got an unexpected result. What should they do next?

   a. They should continue experimenting until they discover the experiment that proves their prediction.

   b. They should repeat the same experiment until they get the result they predicted.

   c. They should repeat the experiment and if they get the same result, they should think about why their prediction was wrong.

   d. They should stop doing any more experiments since their prediction was wrong.

16. There are many goals in doing scientific work, but the main goal of science is

   a. to do experiments.

   b. to discover all of the facts about nature.

   c. to make new inventions.

   d. to build better explanations of things in nature

17. Margaret saw some fuzzy green stuff on a rock she found in the park. When she asked her science teacher about it, the teacher said that it was a kind of plant called moss. Margaret was curious about why the moss grew on the rock. She remembered that the rock was in a dark, damp place. In order to study why moss grows on rocks, what should Margaret do?

Number the following items 1 through 4 in the order you think Margaret should do them.

___ Do an experiment that changes one variable but controls all the other variables.

___ Come up with a hypothesis (a testable idea) about how either moisture or light affects the growth of moss.

___ Look at the result of her experiment and think about what it means.

___ Think about what she already knows about light, moisture and the growth of other plants.

18. Margaret has an idea that moisture is necessary for moss to grow, so she designs an experiment where she puts one rock on a pot of wet dirt and another rock on a pot of dry dirt. Before she begins her experiment, should Margaret think about what the results might be?

a. Yes, because she wants to prove that she has the right plan for her experiment.

b. No, because she cannot possibly predict what the results might be.

c. Yes, because she must think about what different possible results will mean for her idea.

d. No, because she must wait until after she gets the results to think about what different possible results will mean for her idea.

19. A scientist usually makes the most progress in developing an idea by

a. making predictions based on the idea and testing those predictions.

b. rejecting the ideas used by other scientists in their experiments.

c. repeating the same experiment until the expected result happens.

d. trying new things and seeing what happens.

20. The main reason for doing experiments in science is:

a. to come up with new cures or inventions.

b. to try something new because it has never been done before.

c. to explore and test the ideas of scientists.

d. to see what happens when things are mixed together.

Questions 21 - 24 refer to the following situation.

Fred's family had a farm where spotted cows were raised. He was responsible for taking care of a calf (a baby cow). He noticed that the calf went to its own mother for milk. He wondered how the calf could tell its mother from all the other spotted cows on the farm. Fred had the following idea to explain how a calf could tell which cow was its mother.

IDEA: A calf recognizes its mother by the particular pattern of spots the mother has.

To test his idea, Fred decided to watch one particular calf at feeding times. Some of the observations he made support his idea and some observations do not support his idea.

For each of the following observations, circle whether you think the observation SUPPORTS or DOES NOT SUPPORT Fred's IDEA.

| | | |
|---|---|---|
| 21. SUPPORTS | DOES NOT SUPPORT | The calf did not go to a cow that had a different pattern of spots than the calf's mother. |
| 22. SUPPORTS | DOES NOT SUPPORT | Fred painted a picture of a cow that had the same pattern of spots as the mother cow and the calf went to it for milk. |
| 23. SUPPORTS | DOES NOT SUPPORT | The calf went looking for milk from a cow that had no spots. |
| 24. SUPPORTS | DOES NOT SUPPORT | Fred added lots of extra spots to the mother cow using washable paint, and the calf did not go to the mother for milk. |

Questions 21 - 24 refer to the following situation.

Y

Jeff's family had a farm where horses were raised. He was responsible for taking care of a colt (a baby horse). He noticed the colt went to its own mother at feeding time. He wondered how the colt could tell its mother from all the other horses on the farm. Jeff thought about this for awhile and had the following idea to explain how a colt could tell which horse was its mother.

> IDEA: A colt recognizes its mother by the particular neigh (sound) the mother makes.

To test his idea, Jeff decided to watch one particular colt at feeding times. Some of the observations he made support his idea and some observations do not support his idea.

For each of the following observations, circle whether you think the observation SUPPORTS or DOES NOT SUPPORT Jeff's IDEA.

| | | | |
|---|---|---|---|
| 21. | SUPPORTS | DOES NOT SUPPORT | The colt was on the other side of the pasture from its mother. When its mother neighed, the colt went over to her for milk. |
| 22. | SUPPORTS | DOES NOT SUPPORT | Jeff hid the mother horse behind a barn where the colt could not see her. When she neighed the colt did not go to her. |
| 23. | SUPPORTS | DOES NOT SUPPORT | Jeff went inside the barn and played a tape-recording of the mother's neighing. The colt went to the barn. |
| 24. | SUPPORTS | DOES NOT SUPPORT | A horse that looked just like the colt's mother neighed and the colt did not go to her. |

Appendix 5

Study 2 Revised Clinical Interview

WORDS TO UNPACK DURING THE INTERVIEW (What do you mean by __?):

| | | |
|---|---|---|
| Answer | Helps | Theory |
| Conclusion | Learn | Truth |
| Discover | Procedure | Try Again |
| Equipment | Proof | Try Out |
| Explanation | Test | Understand |

---

INTRODUCTORY QUESTIONS:

1) What do you think science is all about?
2) What do you think the goal of science is?
3) Which statement do you think is a better description of the goals of science?
    - I) The goal of science is to discover new things in the world and the universe.
    - II) The goal of science is to build a better understanding of the world around us.

   Why? Can you give me some examples (of new things, or the kinds of things we try to understand)?
4) How do you think a scientist does this work?

---

I. IDEAS

1) Where do scientists get their ideas?
2) What kinds of ideas do scientists have?
3) What are scientists' ideas about?

4a) Do scientists do anything with their ideas? What do they do with them?
If TEST then:
- 4b) How do scientists test their ideas?
- 4c) What happens to the ideas once they've been tested?

5) Is there a relationship between a scientist's ideas and the rest of the work a scientist does? What is the relationship?
6) Do scientists change their ideas? Why (when) or why not?

---

II. HYPOTHESIS

1) What is a hypothesis?
2) Where does a scientist get a hypothesis?
3) Is there a relationship between a scientist's hypotheses and the rest of the work a scientist does? What is the relationship?

---

III. EXPERIMENT

1) What is an experiment? [UNPACK THE ANSWER]

2a) Why do scientists do experiments?
If TO TEST IDEAS then:
- 2b) How does the test tell the scientist something about the idea?

3) How does a scientist decide what experiment to do?
4) Is there a relationship between the mental work a scientist does and the experiments a scientist does? What is the relationship?

---

IV. RESULTS

1) What happens when a scientist is testing his/her ideas, and gets a different result from the one he/she expected? [UNPACK THE ANSWER]