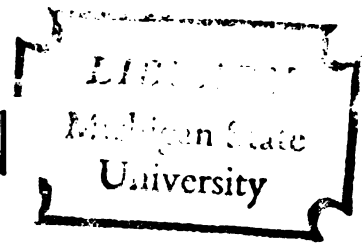


THE DEVELOPMENT AND USE OF A GROUP PROCESS
TEST FOR SELECTED PROCESSES OF THE
SCIENCE CURRICULUM IMPROVEMENT STUDY

Thesis for the Degree of Ph. D.
MICHIGAN STATE UNIVERSITY
JOSEPH WILLIAM RILEY
1972



This is to certify that the

thesis entitled

THE DEVELOPMENT AND USE OF A GROUP PROCESS TEST FOR
SELECTED PROCESSES OF THE SCIENCE CURRICULUM IMPROVE-
MENT STUDY

presented by

Joseph William Riley

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Education

Glen A. Berkhimer

Major professor

Date October 14, 1972

ABSTRACT

THE DEVELOPMENT AND USE OF A GROUP PROCESS TEST FOR SELECTED PROCESSES OF THE SCIENCE CURRICULUM IMPROVEMENT STUDY

By

Joseph William Riley

The purpose of this study is twofold: (1) To develop a paper and pencil science process test for fifth grade students based on some of the stated or implied goals for grades 3, 4, and 5 of the Science Curriculum Improvement Study; and (2) To use this test and other tests to compare fifth grade students who have been in the SCIS program for five consecutive years with similar fifth grade students who have been enrolled in traditional textbook series science courses.

Items were written to satisfy the test specification grid, which included the main content areas specified in grades 3, 4, and 5 of the SCIS program, and four process areas of identifying and controlling variables, interpreting data, predicting, and inferring.

Validity for the test was provided by a panel of science educators, and correlations with the Science Test (Form 4A) of the Sequential Test of Educational Progress

(STEP) Series II. Reliability was determined by the Kuder-Richardson Formula #20.

Two pilot testing programs were held before the final form of the test was developed. The first tryout of the Test of Science Inquiry Skills (TSIS) involved 902 fifth grade students from the school districts of Van Dyke, Livonia, and East Lansing, Michigan. This first tryout was to clarify ambiguities in the instructions, format, and test items. Also, it provided data for calculating test item difficulty and discrimination indices.

The second pilot test program included 185 fifth grade students from the school districts of Waterford, Van Dyke, and East Lansing, Michigan. The purpose of the second tryout program was to gather additional item analyses data. Many items on the revised preliminary form were altered on the basis of data gathered after the first pilot test. Item analyses data obtained from the second pilot test was the major determining factor for selecting items to be included in the final version of the TSIS.

The final form of the test, which contained 50 items, was administered to 310 fifth grade students in 12 classrooms in the school districts of DeWitt, Grand Ledge, and Perry, Michigan. Children in these schools had been in the SCIS program for five years. A Control group made up of 191 children in seven classrooms was selected from schools similar to the SCIS schools.

The final version of the TSIS was administered to the above-mentioned sample of SCIS and Control students. Significant data about the test includes the following for each population:

- (1) SCIS Population: Item difficulty (percentage of the total group that got the item wrong) of 46, item discrimination of 49, standard error of 3.0810, and Kuder-Richardson Formula #20 reliability coefficient of 0.9033;
- (2) Control Population: Item difficulty of 39, item discrimination of 43, standard error of 3.1017, and Kuder-Richardson Formula #20 reliability coefficient of 0.8726.

The second purpose of this study was to use the TSIS and other tests to compare fifth grade students who have been in the SCIS program for five consecutive years with similar fifth grade students who have been enrolled in traditional textbook series science courses.

Twelve SCIS and seven Control classrooms were compared using scores from the Test of Science Inquiry Skills, Science and Scientists Attitude Inventory, Sequential Test of Educational Progress (Series II, Form 4A Science Test), and the composite achievement score of the 1970-71 State of Michigan Assessment Examination. Results were analyzed using a repeated measures design which indicated there was no main effect or significant difference between SCIS and

Control classrooms on any of the four measures used in the study.

Evidence from this study indicates that the paper and pencil test format can function effectively as a group process test for fifth grade students, and that this group test provides a means of evaluating the students' ability to utilize process skills. Students at the fifth grade level had no difficulty in completing and marking machine-scorable answer sheets. Also, they had no difficulty with the test itself; they could follow the questions and relate the test items to specific diagrams. Diagrams, charts, graphs, and pictures can thus be used to effectively communicate problem situations to fifth graders.

One of the important outcomes of this study concerns its implications for future research. Suggestions for future studies include the following:

- (1) Teachers teaching a new program for the first time may not be as effective as teachers teaching a program for the third or fourth consecutive year. Thus, a longitudinal study of the SCIS teachers used in this study three or four years hence may prove valuable.
- (2) Another suggestion for future research would be to utilize individual competency measures, such as the AAAS Competency Measures to compare rural with urban students, and students with and without the new

process-centered science programs. This would shed some light on what effect a child's background, personal experiences, and maturity have on his ability to perform intellectual skills such as problem solving.

- (3) A problem-solving test should be developed that would measure various levels in the process skills hierarchy. This would not only reveal more information about the student, but could also be used as a diagnostic tool to further guide the student in his learning experience.

THE DEVELOPMENT AND USE OF A GROUP PROCESS TEST
FOR SELECTED PROCESSES OF THE SCIENCE
CURRICULUM IMPROVEMENT STUDY

By

Joseph William Riley

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

College of Education

1972

G77980

Copyright by
JOSEPH WILLIAM RILEY
1972

ACKNOWLEDGMENTS

I wish to express my sincere appreciation to the many people who have given generously of their time and talents throughout the duration of this study. I wish to thank Dr. Glenn D. Berkheimer, my advisor and doctoral committee chairman, whose competent guidance, encouragement, and valuable suggestions throughout my doctoral studies have made this a most meaningful experience.

Sincerest thanks also go to Dr. Clarence H. Nelson for his endless patience and guidance throughout the development and field testing of the Test of Science Inquiry Skills. His expertise and encouragement are greatly appreciated.

I am grateful also to Drs. Julian R. Brandou, William Mehrens, and Frank Peabody, members of my doctoral committee, who have also given me much advice, encouragement, and support.

For their cooperation and courtesy during the entire trial testing programs, appreciation is extended to the Science and Mathematics Teaching Center at Michigan State University and the many administrators, teachers, and pupils who participated in the study. I am also grateful to my employer, Van Dyke Public Schools, for the time, encouragement, and resources extended to complete this study.

Finally, I am grateful to my dear wife, Carole, for her constant support, patience, encouragement, help with the drawings used in the test items, and the typing and editing of materials as they were written.

TABLE OF CONTENTS

	Page
LIST OF TABLES	vii
LIST OF FIGURES.	ix
LIST OF APPENDICES	x
Chapter	
I. INTRODUCTION	1
Purpose of the Study	1
Background of the Study.	1
Importance of the Study.	3
Test Development	6
Test Format.	7
Definition of Terms.	7
Hypotheses	8
Procedures in Data Collection.	9
Sample	9
Instruments Used in the Study.	9
Assumptions.	10
Limitations of the Study	11
Organization of the Thesis	11
II. REVIEW OF RELATED LITERATURE	13
Recent Changes in Elementary	
Science Education.	13
The Process Approach to Teaching Science . .	16
Evaluation of New Elementary Science	
Programs	23
Research Studies Involving SCIS.	28
Descriptive Feedback	28
Experimental	30
Development of Instruments to Measure	
Process Skills	34
Methods Used in Test Construction.	49
Preliminary Considerations	49
Item Writing	51
Multiple Choice Items.	52
Experimental Tryout of Test Materials. . .	55
Reliability, Validity, Item Difficulty	
and Discrimination	57

	Page
Studies Which Compare Problem Solving With Achievement and Attitude.	57
Summary.	65
III. DESCRIPTION OF THE STUDY	66
The General Procedure.	66
Identification of Science Processes.	67
Identification of Content Areas.	70
Writing of Test Items.	74
Validation of Test Items	76
Instructions and Answer Sheets	76
First Pilot Test	77
Sample	77
Purpose.	78
Procedure.	78
Results.	79
Second Pilot Test.	80
Sample	80
Purpose.	81
Procedure.	81
Results.	81
Construction of Final Test	81
Length of Test	81
Final Identification of Processes.	82
The Sample	84
Use of Other Instruments in the Study.	89
Science and Scientists Attitude Inventory (SASAI).	89
State of Michigan Assessment Composite Achievement Score.	89
Sequential Test of Educational Progress (STEP) Series II	91
Design of Study.	92
Final Testing Program.	93
Contact With Schools	93
Administration of Tests.	95
Item Analyses Used in the Study.	95
Item Difficulty.	95
Item Discrimination.	95
Reliability.	96
Standard Error of Measurement.	96
Summary.	98
IV. ANALYSIS AND INTERPRETATION OF THE DATA.	102
Test Item Analyses	103
Standardization.	106
Reliability.	106
Validity	109

	Page
Hypotheses	118
Design of the Study.	119
Repeated Measures.	121
Results.	124
Summary.	125
V. SUMMARY AND CONCLUSIONS.	126
Development of the Test.	126
The Study.	128
Limitations.	128
Findings and Conclusions	129
Implications for Education	133
Suggestions for Future Research.	135
BIBLIOGRAPHY	138
APPENDICES	147

LIST OF TABLES

Table	Page
1. Comparison of SCIS and Control Schools	88
2. Years of Students' SCIS Experience in Experimental Classrooms.	94
3. Grouping of Test Items Into Subtests on the Basis of Process	99
4. Final Test Specification Grid.	100
5. Summary Data on Items for Test of Science Inquiry Skills (SCIS Population)	104
6. Summaries of Reliability for SCIS and Control Populations; and for Subtests With SCIS Population Only.	108
7. Summary Data for Subtest Identifying and Controlling Variables.	110
8. Summary Data for Subtest Interpreting Data . . .	111
9. Summary Data for Subtest Predicting.	112
10. Summary data for Subtest Inferring	113
11. Summary of Classroom and School Means Comparison by SCIS and Control Schools	116
12. Summary of the Means on the Four Tests by Classroom	122
13. Results of Repeated Measures by the Use of Analysis of Variance.	123
14. TSIS Analyses Data	127
B1. TSIS Raw Score Distributions--SCIS Population. .	178
B2. TSIS Summary Data--SCIS Population	179
B3. TSIS Raw Score Distributions--Control Population	180

Table	Page
B4. TSIS Summary Data--Control Population.	181
B5. STEP Raw Score Distributions--SCIS Population. .	182
B6. STEP Raw Score Distributions-- Control Population	183
B7. SASAI Raw Score Distributions--SCIS Population .	184
B8. SASAI Raw Score Distributions-- Control Population	185
C1. Summary Data on Items for TSIS, Form A, Second Pilot Test.	216
C2. Summary Data on Items for TSIS, Form B, Second Pilot Test.	244

LIST OF FIGURES

Figure	Page
1. Final Edition Grade Level Concepts of SCIS Program.	4
2. Preliminary Test Specification Grid	75
3. Final Test Specification Grid	83
4. General Design of Study	93
5. Sample of Item Analyses Data Sheet.	97
6. Correlation Coefficients Between Test Scores. .	117
7. General Design of Study	120
8. Summary of Treatment by Measures Data	124
C1. Preliminary Test Specification Grid, Form A, Second Pilot Test	218
C2. Preliminary Test Specification Grid, Form B, Second Pilot Test	246

LIST OF APPENDICES

Appendix	Page
A. A TEST OF SCIENCE INQUIRY SKILLS AND SAMPLE ANSWER SHEET.	148
B. TEST OF SCIENCE INQUIRY SKILLS, SEQUENTIAL TEST OF EDUCATIONAL PROGRESS, AND SCIENCE AND SCIENTISTS ATTITUDE INVENTORY RAW SCORE DISTRIBUTIONS FOR SCIS AND CONTROL POPULATIONS.	177
C. PRELIMINARY FORMS A AND B, TEST OF SCIENCE INQUIRY SKILLS	186
D. OVERVIEW AND INSTRUCTIONS GIVEN TO PARTICIPATING PRINCIPALS AND TEACHERS.	247
E. LIST OF PARTICIPATING SCHOOLS.	256

CHAPTER I

INTRODUCTION

Purpose of the Study

The purpose of this study is twofold:

(1) To develop a paper and pencil science process test for fifth grade students based on some of the stated or implied goals for grades 3, 4, and 5 of the Science Curriculum Improvement Study (SCIS);¹

(2) To use this test and other tests to compare fifth grade students who have been in the SCIS program for five consecutive years with similar fifth grade students who have been enrolled in traditional textbook series science courses.

Background of the Study

In 1962, a nationwide Commission on Science Instruction was established to assist in exploring new approaches for teaching science to children. This commission, established by the American Association for the Advancement of Science (AAAS), is now called the Commission on Science Education. Nearly a dozen curriculum projects have been initiated since 1962, each with a special point of view, but all with the idea of presenting science in its true character

¹SCIS will be used throughout this study as the abbreviation for Science Curriculum Improvement Study.

and with educational goals that are suited to modern times. These projects agree that science learning requires active involvement of the children, who perform experiments, make observations, and draw conclusions.²

One of these new elementary projects, the SCIS, was established in 1962 by Robert Karplus, a professor of theoretical physics at the University of California, Berkeley, and was funded by the National Science Foundation. The curriculum uses a materials-centered approach so that the children come in direct physical contact with natural phenomena and have the freedom to explore and thus discover new concepts themselves.

A central objective of SCIS is the development of scientific literacy. Scientific literacy is described as sufficient knowledge and understanding of the fundamental concepts of both the biological and physical sciences for effective participation in twentieth century life. A second implication of scientific literacy is the development of a free and inquisitive attitude and the use of rational procedures for decision making. Each unit of the SCIS program presents activities which lead to the understanding of

²Paul DeHart Hurd, "Elementary Science for the 1970's in Perspective," Developments in Elementary School Science (Washington, D. C.: American Association for the Advancement of Science, Misc. Pub. 70-18, 1970), p. 7.

important scientific concepts and to the development of process skills.³

Figure 1 shows all six levels of the SCIS program,⁴ along with the concepts introduced in each unit.

Importance of the Study

There are many people who claim that the children educated in the sixties are far better educated than those in the fifties, the forties, and the thirties. The truth of the matter, however, is that we do not know. Few measures were taken to show whether the children in the sixties were better educated or more poorly educated than were their predecessors.⁵

There is a real need for more information concerning the relative strengths and weaknesses of the new lab-centered elementary science programs as compared with the more traditional textbook-centered programs.⁶ Questions arise which

³J. David Lockard, ed., Seventh Report of the International Clearinghouse on Science and Mathematics Curriculum Developments, 1970 (College Park, Maryland: University of Maryland and AAAS, 1970), pp. 532-44.

⁴Science Curriculum Improvement Study, SCIS Newsletter, ed. by Suzanne Stewart, 17 (1970), 3.

⁵Gilbert R. Austin, "Educational Accountability--Hallmark of the 1970's," The Science Teacher, XXXVIII (April, 1971), 26-28.

⁶Wayne W. Welch, "Curriculum Evaluation," Review of Educational Research, XXXIX, 4 (1969), 429-43; J. Myron Atkin, "Some Evaluation Problems in a Course Content Improvement Project," Journal of Research in Science Teaching, I, 1 (1963), 129-32; Leopold E. Klopfer, "Effectiveness and Effects of ESSP Astronomy Materials--An Illustrative Study of Evaluation in a Curriculum Development Project," Journal of Research in Science Teaching, VI, 1 (1969), 64-75.

Grade 1			
ORGANISMS		MATERIAL OBJECTS	
organism	habitat	object	serial ordering
birth	food web	property	change evidence
death	detritus	material	
Grade 2			
LIFE CYCLES		INTERACTION AND SYSTEMS	
growth	biotic potential	interaction	system
development	generation	evidence of	interaction-at-
life cycle	plant and animal	interaction	a-distance
Grade 3			
POPULATIONS		SUBSYSTEMS AND VARIABLES	
population	plant eater	subsystem	temperature
predator	animal eater	histogram	variable
prey	food chain	solution	
community	food web		
Grade 4			
ENVIRONMENTS		RELATIVE POSITION AND MOTION	
environment	range	reference object	reference frame
environmental	optimum range	relative position	polar coordi-
factor		relative motion	nates
			rectangular
			coordinates
Grade 5			
COMMUNITIES		ENERGY SOURCES	
photosynthesis	producer	energy transfer	energy source
food pyramid	consumer	energy chain	energy receiver
community	decomposer		
Grade 6			
ECOSYSTEMS		MODELS: ELECTRIC AND MAGNETIC INTERACTION	
ecosystem	oxygen-carbon	model	electrochemical
water cycle	dioxide cycle	electric current	cell
	pollution	magnetic field	series/parallel
		electrode	circuits

Figure 1.--Final edition grade level concepts of SCIS program.

concern the effect of SCIS itself as well as how it compares in effectiveness to other available materials for teaching elementary science.⁷ For this reason, instrument development in elementary science needs more careful attention, and more research needs to be done in this area.⁸

Current trends in science education have created a need for instruments with which to measure students' progress toward previously defined objectives within the newer framework of process- and lab-centered programs. A search of the literature revealed that, with the exception of the AAAS competency measures, no group instrument was available with which to measure fifth grade students' ability to use science processes. This study, which includes the development of the Test of Science Inquiry Skills (TSIS)⁹ is an attempt to partially fill this need. The instrument is for measuring progress toward behavioral objectives which stress science

⁷Lockard, Seventh Report on Science and Mathematics, pp. 532-44; Henry H. Walbesser, "Science Curriculum Evaluation: Observations on a Position," The Science Teacher, XXXIII (February, 1966), 34-39.

⁸David P. Butts and Howard L. Jones, "The Development of the TAB Science Test," Science Education, LI (December, 1967), 463-73; Wayne W. Welch and Milton O. Pella, "The Development of an Instrument for Inventorying Knowledge of the Processes of Science," Journal of Research in Science Teaching, V, 1 (1967), 64-68; Gregor A. Ramsey and Robert W. Howe, "An Analysis of Research Related to Instructional Procedures in Elementary School Science," Science and Children, VI (April, 1969), 25-36.

⁹TSIS will be used throughout this study as the abbreviation for Test of Science Inquiry Skills.

processes. It should enable teachers to prepare more reliable and valid evaluation of their students' achievement in these areas which are now being emphasized in elementary science.

Data gathered from the TSIS developed for this study could be useful in considering the following questions regarding SCIS and other process-centered science programs:

- (1) Does the program meet some of its stated objectives? and
- (2) Do students attain these same objectives in other elementary science programs which are not process centered?

Further, research involving evaluative instruments may be of practical value to school administrators in helping them make more objective decisions regarding their courses of instruction. Many elementary schools are under increased pressure to upgrade and/or change the science curriculum. However, school administrators and teachers find themselves in a position of making important decisions with little factual information as to what to expect of their students if they adopt one of the new programs. Utilization of instruments such as the one developed in this study may thus be helpful in this regard.

Test Development

In light of the above-mentioned need for a group process instrument, a major purpose of this study was to develop a paper and pencil test to measure process skills of fifth grade science students. Acceptable guidelines for

test development procedures, as established in the literature, are reviewed in Chapter II. A more detailed description of the test development procedure follows in Chapter III.

Test Format

The test is a 50-item multiple choice paper and pencil test. It is designed for fifth grade students and can be easily administered in 60 minutes to an entire class by an individual classroom teacher. Most of the items present a situation by a drawing, diagram, or chart. A series of questions is then asked that can be answered based on the given information.

Definition of Terms

The following terms and definitions are used throughout this study:

AAAS--American Association for the Advancement of Science

Item difficulty--proportion of the total group that got the item wrong

Item discrimination--difference between the proportion of the upper 27% of the group who got the item correct and the proportion of the lower 27% who got the item correct

KR #20--Kuder Richardson Formula #20

SASAI--Science and Scientists Attitude Inventory

SCIS--Science Curriculum Improvement Study

STEP--Sequential Test of Educational Progress

TSIS--Test of Science Inquiry Skills

Hypotheses

As previously mentioned, the purpose of this study is to develop a process instrument and to use this and other instruments to estimate to what degree the SCIS program has met some of its stated goals. The most specific hypotheses to be tested are as follows:

Ho₁: There is no significant difference between classrooms receiving science instruction with SCIS and classrooms not receiving SCIS instruction as measured by the Science and Scientists Attitude Inventory (SASAI), the Test of Science Inquiry Skills (TSIS), the Sequential Test of Educational Progress (STEP), and the State of Michigan Assessment.

Ho₂: On the Science and Scientists Attitude Inventory, the Test of Science Inquiry Skills, the Sequential Test of Educational Progress, and the State of Michigan Assessment, there will be no significant interaction between treatment and measures.

Ho₃: There is no significant difference between classrooms receiving SCIS science instruction and Control classrooms when comparing their achievement as measured by the State of Michigan Assessment composite achievement score.

Ho₄: There is no significant difference between classrooms receiving SCIS instruction and Control classrooms when comparing their ability to solve science problems as measured by the Test of Science Inquiry Skills.

Ho₅: There is no significant difference between classrooms receiving SCIS instruction and Control classrooms when comparing their ability to recall factual science information as measured by the Sequential Test of Educational Progress (STEP).

Ho₆: There is no significant difference between classrooms receiving SCIS instruction and Control classrooms when comparing their attitudes toward science and scientists as measured by the Science and Scientists Attitude Inventory (SASAI).

Procedures in Data Collection

The following section outlines the general procedure followed in data collection. A more detailed description follows in Chapter III.

Sample

Twelve SCIS classrooms in three school districts were designated as the experimental group in the final testing program. Seven Control classrooms in two school districts were selected based on the similarity on measured parameters on the 1970-71 State of Michigan Assessment Examination to the SCIS schools.

Instruments Used in the Study

Besides the Test of Science Inquiry Skills (TSIS), three other instruments, the Science and Scientists Attitude Inventory (SASAI), the Sequential Test of Educational Progress

(STEP), and the State of Michigan Assessment composite achievement scores for these students when they were in the fourth grade, were utilized in the study to obtain a more balanced assessment of the effect of the SCIS program.

Assumptions

The Test of Science Inquiry Skills is based on the following assumptions:

- (1) The four processes of identifying and controlling variables, interpreting data, predicting, and inferring, are representative members of the basis science processes taught in the SCIS program in grades 3, 4, and 5.
- (2) The content areas identified are representative of the content taught in the SCIS program.
- (3) A carefully selected committee of science educators can provide test item validity.
- (4) The Test of Science Inquiry Skills has content validity because it is composed of valid items.
- (5) The reliability of the Test of Science Inquiry Skills is acceptable if the correlation coefficient as determined by the Kuder Richardson Formula #20 is equal to or greater than 70.
- (6) Acceptable educational measurements have been developed using the procedures followed in this study.

- (7) Students who participated in this study were representative of other students in learning abilities and science process achievement potential.

Limitations of the Study

The following questions point out some of the possible limitations of the study:

- (1) Can specific process skills be measured by paper and pencil tests?
- (2) Can diagrams, charts, pictures, etc. communicate what they are intended to communicate?
- (3) Will children in the testing program take the tests seriously?
- (4) How much of a factor is the student's reading ability in taking these tests?
- (5) It is difficult, if not impossible, to maintain intact classes from year to year. Also, students are not always randomly assigned to classrooms. In this study, the unit is the classroom. How much of an effect will these variables have on the study?

Organization of the Thesis

The general organization of the thesis is as follows: Chapter II reviews the literature relative to changes in elementary science programs, the need for evaluation of the new process-centered programs, test construction, and the development of an appropriate evaluative instrument. Chapter III describes the procedures used in developing and

pilot-testing the TSIS. The use of the TSIS, along with three other instruments used in the study, is also discussed. Chapter IV includes the analyses and interpretation of the data obtained from the study, and Chapter V reports the findings and conclusions of the study.

CHAPTER II

REVIEW OF RELATED LITERATURE

Several aspects of science education directly related to this study are: changes in elementary science programs, the need for evaluation of these new process-centered programs, test construction, and the development of an appropriate evaluative instrument.

This chapter is organized into the following sections: (1) Recent changes in elementary science education; (2) The process approach to teaching elementary science; (3) Evaluation of new elementary science programs; (4) Research studies involving the Science Curriculum Improvement Study (SCIS); (5) Development of instruments to measure process skills; (6) Methods used in test construction; (7) Studies which compare problem-solving or process skills with achievement and attitude; and (8) Summary.

Recent Changes in Elementary Science Education

An analysis by Mallinson¹ in 1961 of the then current status of elementary school science curriculums pointed to confusion in sequencing of science topics; disagreement

¹Jacqueline B. Mallinson, "The Current Status of Science Education in the Elementary School," School Science and Mathematics, LXI (April, 1961), 252-70.

about time allocation afforded science in selected school districts; and that all accepted methods of teaching seemed effective when properly used. Her study represented one phase of a comprehensive study undertaken by a committee of the American Association for the Advancement of Science (AAAS) and was designed to review the status of elementary school science and to formulate an extensive plan for certain improvements.

Recommendations as a result of these studies and regional conferences attended by scientists, psychologists, teachers, science supervisors, and science educators included the following points: (1) Science should be a basic part of general education, starting in the primary grades for all students; (2) There must be clear progression from grade to grade; (3) Science teaching should stress the "spirit of discovery characteristic of science itself"; and (4) Preparation of new instructional materials requires the combined efforts of scientists, classroom teachers, and specialists in learning and teacher preparation. The report advocated that major attention be given to the teaching of cognitive processes integral to the scientific enterprise. A massive, coordinated attack on the improvement of science teaching was recommended.²

²J. Myron Atkin, "Science in the Elementary School," Review of Educational Research, XXXIV, 3 (1964), 267.

Ipsen described elementary science of the early 1960's as a program that contained almost nothing that could be called a study of scientific activity.³ The question of how scientific information is obtained was almost wholly neglected. Further, there was little in these programs that a scientist could recognize as being science.⁴

These are but a few of the numerous references to elementary science in the early 1960's--a period well-documented to be one in which programs were designed to do little more than give children scientific and technological information.

After the AAAS conferences released their report in the spring of 1961, several study groups were set up along the lines suggested. The AAAS established a Commission on Science Education, which supervised the creation of new curriculum materials. The Commission also set up a clearing house for information concerning curriculum developments in elementary school science and mathematics at the University of Maryland. Nearly a dozen curriculum projects have been initiated since 1962, each with a special point of view, but all with the idea of presenting science in its true character and with educational goals that are suited to modern

³D. C. Ipsen, Issues in Elementary Science (Washington, D. C.: National Science Teachers Association, 1970), p. 3.

⁴Paul DeHart Hurd, "Elementary School Science for the 1970's in Perspective," Developments in Elementary School Science (Washington, D. C.: American Association for the Advancement of Science, Misc. Pub. 70-18, 1970), p. 3.

times. The writers of these projects agree that science learning requires active involvement of the children, who perform experiments, make observations, and draw conclusions.⁵

The period of the early 1960's was one of broad-ranging innovation in curriculum and teaching styles; a threshold period in which major shifts occurred in some of the fundamental aims of science instruction.⁶

The Process Approach to Teaching Science

A look at the new science programs indicates an activity-oriented approach toward the development of some fundamental notions, activities which are carefully sequenced and which have a great deal of involvement of the pupils with objects and ideas. Presently, according to Fischler,⁷ there are three schools of thought regarding elementary science. There is a process approach in which the processes are the major emphasis. There is a conceptual approach in which the conceptual scheme receives the major emphasis. Third, there is an approach which takes a unit, sequences the processes and sequences the discrepant events (inconsistencies in a child's conceptual background) in a logical pattern.

⁵Robert Karplus and Herbert D. Thier, A New Look at Elementary Science (Chicago: Rand McNally & Company, 1967), p. 3.

⁶Atkin, "Science in the Elementary School," p. 270.

⁷Abraham Fischler, "Implications of Structure for Elementary Science," Science Education, LII, 3 (1968), 278-79.

Statements of philosophy or objectives underlying these new elementary science courses include claims for "scientific literacy," "learning by discovery," and "teaching the process of science." To enable students to understand science and scientists better, teachers must give continual and systematic emphasis to the process of science during their instructional activity.⁸

Why should the process of science become the organizing emphasis for science instruction in the elementary grades? Gagné outlined five advantages of the process emphasis:

- (1) Processes like observing, classifying, measuring, inferring, and so on, represent some basic ways of dealing with the natural environment which are not as systematically taught in the elementary grades as they ought to be.
- (2) Processes, as they are built up from the relatively simple forms of observing and classifying to the relatively complex forms like formulating operational definitions and interpreting data are intellectual strategies which are believed to be valuable.
- (3) As intellectual skills, processes have the obvious potential for high degrees of transfer to many activities.
- (4) The intellectual skills which processes represent presumably do not have to be learned as the student progresses from grade two to grade eight or grade twelve. Fundamentally, particular examples of observing and measuring are the same processes whether they are engaged in by a seven-year-old, an eighteen-year-old, or by a mature scientist.
- (5) Processes are general to all sciences, and thus make the problem of selection an easy one, as well

⁸Marshall A. Nay and Associates, "A Process Approach to Teaching Science," Science Education, LV, 2 (1971), 197.

as one that is likely to be beneficial regardless of later choices of courses of study by the student.⁹

The National Science Teachers Association stated its position on "School Science Education for the 1970's" and includes:

The major goal of science education is to develop scientifically literate and personally concerned individuals with a high competence for rational thought and action. This choice of goals is based on the belief that achieving scientific literacy involves the development of attitudes, process skills and concepts, necessary to meet the more general goals of all education.¹⁰

The AAAS Commission on Science Education formulated a statement of its objectives for science education in the elementary school to serve as a guidepost. This statement was prepared by Kessen and includes the following:

Science is best taught as a procedure of enquiry. . . . Science is more than a body of facts, a collection of principles, and a set of machines for measurement; it is a structured and directed way of asking and answering questions. . . . The procedures of scientific enquiry, learned not as a canon of rules but as ways of finding answers, can be applied without limit. . . . The use of laboratory techniques--especially the experiment--deserves special attention.¹¹

With all this emphasis given the "process approach," it seems appropriate to ask: What is the psychological

⁹Robert M. Gagné, "Why the 'Process' Approach for a Modern Curriculum?" Educational Products Information Exchange, I, 8 (1968), 11-12.

¹⁰National Science Teachers Association, "School Science Education for the 1970's," The Science Teacher, XXXVIII (November, 1971), 46-51.

¹¹William Kessen, "Statement of Purposes and Objectives of Science Education in School," The Psychological Bases of Science--A Process Approach (Washington, D. C.: American Association for the Advancement of Science, Misc. Pub. 65-8, 1965), pp. vi-xi.

basis of the process approach and how is it defined in the literature?

Writers for the elementary curriculum Science--A Process Approach, under the guidance of the Commission on Science Education of the American Association for the Advancement of Science, used the following working definition:

. . . that processes represent forms of information processing, in this case, on the part of the child. The scientist uses process words, such as classifying, measuring, hypothesizing, to represent, in formal terms, the activities he carries out in pursuing scientific knowledge.¹²

Welch described scientific process as a series of activities or operations performed by the scientist in his attempt to understand nature.¹³ These activities are based on various assumptions and are guided by an awareness of the nature of the outcomes and the ethics and goals of the discipline.

The Science Curriculum Improvement Study, in order to develop an effective program of science instruction within a framework of elementary education, utilized the findings of Piaget, Hunt, Bruner, and Almy. It was concluded that the elementary school years should provide: (1) a diversified program based heavily on concrete manipulative experiences (used guidelines of Piaget); (2) these experiences in

¹²Gagné, "Why the 'Process' Approach," p. 11.

¹³Wayne W. Welch, "The Development of an Instrument for Inventorying Knowledge of the Processes of Science" (unpublished Ph.D. dissertation, University of Wisconsin, 1966).

a context that helps to build a conceptual framework; and (3) a conceptual framework that permits them to perceive phenomena in a more meaningful way, i.e., integrate their influences into generalizations of greater value than the ones they would form if left to their own devices. The curriculum is built around extensive laboratory experiences where the students are involved in exploring new experiences and phenomena. It is referred to as a direct approach to learning.¹⁴

The elementary science program Science--A Process Approach, produced under the sponsorship of the American Association for the Advancement of Science, is the one major effort to develop elementary courses on the basis of a "process" approach. It relies heavily on the psychology of learning developed by Gagné, who outlined the following rationale:

This approach seeks a middle ground between the extremes I have mentioned [the content approach and the creativity view]. . . . Specifically, it rejects the "content approach" idea of learning highly specific facts or principles of any particular science or set of sciences. It substitutes the notion of having children learn generalizable process skills which are behaviorally specific, but which carry the promise of broad transferability across many subject matters. . . . The point of view is that if transferable intellectual processes are to be developed in the child for application to continued learning in sciences, these intellectual skills must be separately identified, and learned, and otherwise nurtured in a highly systematic manner. It is not

¹⁴Barbara S. Thompson and Alan M. Voelker, "Programs for Improving Science Instruction in the Elementary School, Part II: SCIS," Science and Children, VII, 8 (1970), 30-31.

enough to be creative "in general"--one must learn to carry out critical and disciplined thinking in connection with each of the processes of science. One must learn to be thoughtful and inventive about observing, and about predicting, and about manipulating space and time, as well as about generating novel hypotheses.¹⁵

Thus, to understand science, one must understand the process of science, defined by Anderson as ". . . the means of investigation by which the body of knowledge is acquired."¹⁶

The terms "critical thinking," "problem solving," and "inquiry" also appear repeatedly in the literature. All of these encompass the "process" approach because these are some of the processes of science. Nicodemus stated that science is valued through its power in finding answers, but the answers are viewed as a temporary outcome of the main activity of science--inquiry.¹⁷

Dietz and George, for the purposes of their study, defined a problem as an apparent inconsistency in a child's conceptual background.¹⁸ Psychologists call such

¹⁵Robert M. Gagné, "Psychological Issues in Science--A Process Approach," The Psychological Bases of Science--A Process Approach (Washington, D. C.: American Association for the Advancement of Science, Misc. Pub. 65-8, 1965), p. 4.

¹⁶Ronald D. Anderson, "Evaluation in Elementary School Science, Part I: Formulating Objectives," Science and Children, V (September, 1967), 21.

¹⁷Robert B. Nicodemus, "Content and Skill Hierarchies in Elementary Science: An Analysis of ESS Small Things," Journal of Research in Science Teaching, VII, 3 (1970), 173.

¹⁸Maureen A. Dietz and Kenneth D. George, "A Test to Measure Problem Solving Skills in Science of Children in Grades 1, 2, and 3," Journal of Research in Science Teaching, VII, 4 (1970), 342.

inconsistencies, discrepant events. Once a discrepant event is introduced to the child, what does the child have to do?

Piaget conceives the adaptive interaction between organism and environment to involve two complementary processes, corresponding to inner organization and outer adaptation, which he calls assimilation and accommodation. . . assimilation occurs whenever an organism utilizes something from the environment and incorporates it. . . accommodation, the process complementary to assimilation, operates as the variations in environmental circumstances demand coping which modifies existing schemata. [Schemata--observed repeatable and generalized pieces of behavior.]¹⁹

Shulman pointed out the third important principle in Piaget's system:

This is the principle of auto-regulation, or equilibration. Piaget sees the development as a sequence of successive disequilibria followed by adaptations leading to new states of equilibrium. The imbalance can occur because of an ontogenetic change occurring naturally as the organism matures. It can also occur in reaction to an input from the environment. Since disequilibrium is uncomfortable, the child must accommodate to new situations through active modification of his present cognitive structure.²⁰

Fischler pointed out that, in terms of the child, it is important how far the discrepant event is from the child's experience.²¹ If he has never had any experience with a particular event, then he is not even aware there is a problem. On the other hand, if (in the learning situation)

¹⁹Joseph McV. Hunt, Intelligence and Experience (New York: The Ronald Press Company, 1961), pp. 111-12.

²⁰Lee S. Shulman, "Psychology and Mathematics Education," The Sixty-Ninth Yearbook of the National Society for the Study of Education, Part I, ed. by Edward G. Begle (Chicago: University of Chicago Press, 1970), pp. 41-42.

²¹Fischler, "Implications of Structure," p. 279.

we introduce a discrepant event which we think is discrepant, but for which the child can account intellectually, it is no problem either.

Research has shown that there are many components or skills involved in problem solving. Studies by Dressel, Rimoldi, Butts and Jones, and Das²² found that the problem solver will collect data he thinks he needs to solve a problem. He will then use these data and infer, predict, or hypothesize a solution.

Evaluation of New Elementary Science Programs

The need for evaluation of current elementary science programs centering on the processes and activities of science is well-established in the literature.

Welch stated that with a few notable exceptions, the curriculum reform movement in science, characterized by its enlistment of scholars and scientists, has not been subjected to religious research and evaluation.²³ He further suggested that:

²²P. L. Dressel, "Better Tests for Science Classrooms," The Science Teacher, XXVI (September, 1959), 342-45; H. J. A. Rimoldi, Training in Problem Solving (Chicago: Loyola University Press, 1961); David P. Butts and Howard L. Jones, "The Development of the TAB Science Test," Science Education, LI (December, 1967), 463-73; Radha C. Das, "The Problem Development Method," Science Education, XLVIII, 5 (1964), 485-88.

²³Wayne W. Welch, "Curriculum Evaluation," Review of Educational Research, XXXIX, 4 (1969), 429-43.

One way to assess the impact of the new programs is to determine whether the stated goals of the project have been achieved. Is the course modern in perspective, unifying in structure, and selective in content?²⁴

Mayer emphasized that curriculum development without associated evaluative activities cannot make an effective contribution to student learning.²⁵ The cognitive domain seems now well in hand, and we are turning our attention to measuring achievement in the affective domain--a task that promises to be both difficult and rewarding.

Atkin contended that inadequate attention has been given to certain evaluation problems in the broad area of curriculum development, particularly in a period of radical modification of course content such as we see today in mathematics and science, and that a few of the problems are only dimly recognized.²⁶

Klopfer stated that the need for more effective means of evaluating the new science programs is an urgent one.²⁷ Further, means of evaluating behaviors peculiar to these programs have yet to be devised.

²⁴Wayne W. Welch, "The Need for Evaluating National Curriculum Projects," Phi Delta Kappan, XLIX, 9 (1968), 530-32.

²⁵William V. Mayer, "Evaluation and Curriculum Development," BSCS Newsletter, 46 (February, 1972), 2.

²⁶J. Myron Atkin, "Some Evaluation Problems in a Course Content Improvement Project," Journal of Research in Science Teaching, I, 1 (1963), 129.

²⁷Leopold E. Klopfer, "Evaluation of Learning in Science," Handbook on Formative and Summative Evaluation of Student Learning (New York: McGraw-Hill Book Company, 1971), p. 638.

One reason for the lack of proper evaluation of these new programs may be the problem of constructing the proper evaluative instrument. Walbesser and Carter expressed the view that, with involvement being a theme of most, if not all curriculum efforts, it is unlikely that an evaluation of behavior could be restricted to paper and pencil measures of performance and still directly measure the total array of expected behavioral acquisitions.²⁸

Beard expressed an alternative view in her study involving the evaluation of Science--A Process Approach in grades 1, 2, and 3.²⁹ The final editions of this curriculum include individually administerable performance tasks to be used in the primary grades. These competency measures yielded important data about the early edition trials, but their regular use in other classrooms, in her opinion, seems unlikely. Primary teachers do not have the training or the time to objectively administer performance tests to their students at the completion of each of 20 exercises in science. It seems there is a need for alternative evaluation modes. A worthwhile alternate should be more time efficient and yield information that is valid, reliable, and useful.

²⁸Henry H. Walbesser and Heather Carter, "Some Methodological Considerations of Curriculum Evaluation Research," Educational Leadership, XXVI (October, 1968), 55.

²⁹Jean Beard, "The Development of Group Achievement Tests for Two Basic Processes of AAAS Science--A Process Approach," Journal of Research in Science Teaching, VIII, 2 (1971), 179.

In the SCIS Teachers' Guides and Evaluation Supplements, formal evaluation or feedback is an integral part of the program, in which the teacher gathers from students, reactions to specific activities and to the program. In the guide for each unit, the teacher is referred to activities and situations that will yield information regarding the pupil's progress.

Documented evaluation, in which the teacher gathers from student reaction to prescribed activities, and for which he records results, is provided for in the SCIS Evaluation Supplements. Trial editions of these supplements for the units Interaction and Systems³⁰ and Life Cycles³¹ were published in 1971 and attempt to evaluate three aspects of the program: (1) perception of classroom environment, (2) concept/process objectives, and (3) attitudes in science.

The first phase, perception of class environment, answers the questions: "Do your pupils feel that they are active participants in science activities?" and "Do your pupils generally like their science activities?" Student answer sheets utilizing happy, neutral, sad, and angry faces enable a teacher to make comparisons between pupils' perceptions of the science class with her own. By circling

³⁰Science Curriculum Improvement Study, Interaction and Systems: Evaluation Supplement (Berkeley: University of California, June, 1971).

³¹Science Curriculum Improvement Study, Life Cycles: Evaluation Supplement (Berkeley: University of California, June, 1971).

a face, each pupil indicates how he feels about a certain activity.

The second phase (concept/process objectives) answers the question: "How closely is each child attaining the objectives of the unit?" Evaluation materials include materials developed during the particular unit itself; for example, to evaluate attainment of objectives in the unit on Plant Growth and Development, materials would include plants in various stages of development, and seeds. Other materials include pictures of these plants at various stages. Guidelines provided in the supplement suggest that four to six children be evaluated at the same time using these materials and by asking divergent and convergent questions.

The third phase, which attempts to evaluate attitudes in science, is a continual part of the program. As examples, the supplement identifies four areas: (1) curiosity or interest, (2) inventiveness, (3) independent or critical thinking, and (4) persistence. Observations made by the teacher throughout the unit are recorded on the Individual Profile sheet in relation to each of the four areas.

The evaluation supplements are intended to help the teacher document his students' progress, but he is warned against using these records to make comparisons among students. Further, all evaluation activities are carried out

in a relaxed atmosphere, matching as closely as possible the regular SCIS classroom environment.³²

These two trial editions of the Evaluation Supplements were field tested during the school year 1971-72 at the various SCIS Trial Centers throughout the country. Reports of these field testing programs were not available at the time this study was written. It is certain, however, that the result of the field tests will provide the input for necessary revisions.

Research Studies Involving SCIS

Studies involving the SCIS program can be placed in either of two categories: (1) Descriptive Feedback, and (2) Experimental.

Descriptive Feedback

This category includes studies where data have been collected through observation of and/or discussion with teachers, illustrative of what occurs in the classroom. In some instances the teacher is the observer and evaluator, while at other times an outsider fulfills this role.³³

Ness collected data from teachers who were using trial editions of the unit Organisms, and discovered that teachers are an invaluable source of critical analysis of

³²Science Curriculum Improvement Study, SCIS Newsletter, ed. by Suzanne Stewart, No. 22 (1972), p. 3.

³³Thompson and Voelker, "Programs for Improving Science Instruction," p. 8.

materials, and may have been a major resource in this project.³⁴

Flory did an observational study of 28 classrooms and found that a large percentage of the teachers utilized discussion techniques which are contrary to the SCIS philosophy.³⁵ Conclusions from this study suggest that teachers need in-service preparation when utilizing a program of this nature.

Tresmontan conducted a study to determine the needs of teachers as well as pupils and identify ways that SCIS could provide beneficial services for them.³⁶ This study pointed out that content background is important, and could best be provided by in-service training sessions in conjunction with SCIS. Also, it seems that teachers accept and use constructive criticism when they are respected as knowledgeable persons.

These and other similar studies have enabled SCIS to revise and improve existing materials, and improve the training of SCIS teachers--all of which should eventually be reflected in a stronger program.

³⁴Patricia D. Ness, "Organisms Feedback," What Is Curriculum Evaluation?, ed. by Robert Karplus (Berkeley: University of California, 1968), pp. 9-21.

³⁵Donald L. Flory, "Dynamics of Classroom Behavior: An Informal Study of SCIS," What Is Curriculum Evaluation?, ed. by Robert Karplus (Berkeley: University of California, 1968), pp. 22-37.

³⁶Olympia D. Tresmontan, "SCIS in the Classroom, A Sample," What Is Curriculum Evaluation?, ed. by Robert Karplus (Berkeley: University of California, 1968), pp. 38-44.

Experimental

This aspect of SCIS evaluation has been left to researchers. A review of the literature indicates that thus far, these studies have centered on mainly three aspects of the SCIS program: (1) changes in teaching behavior and style, (2) pre- and in-service teacher training programs, and (3) concept development in SCIS children.

Many other studies compiled by Thompson and Voelker³⁷ support the idea that SCIS teachers used a greater number of higher level questions, requiring more thinking, than non-SCIS teachers. Further, comparisons of questioning level before and after SCIS workshops and training programs for teachers reveal that higher level questions were used after these programs. Higher level of questioning involves analysis type questions which elicit greater degree of cognitive skill than does the recall of fact type question.

In regard to teacher preparation, reports from a study conducted in the Magnolia, Arkansas, Public Schools suggest that in-service work is more effective if it is not in large segments. Rather, regular and continuous contacts throughout the year seemed more beneficial to the teachers.³⁸

³⁷Thompson and Voelker, "Programs for Improving Science Instruction," pp. 8-10.

³⁸Final Report on SCIS Activities for the 1968-69 School Year at the Magnolia Public Schools (Magnolia, Arkansas, 1969).

Fischler used SCIS materials as a vehicle for teacher preparation.³⁹ It appears that inherent in SCIS is embodied a philosophy of teacher behavior which is useful in preparing teachers to be self-critical and self-analytical of their behavior.

Many studies have been conducted in the area of concept development with the SCIS materials, each study approaching the problem from a different direction.

Studies by Siegelman and Karplus and by Thier involved concept development in specific units of the SCIS program.⁴⁰ Information from these studies led to revision of the units Relativity and Material Objects. In the Karplus and Siegelman study, the data indicated that of the five objectives tested on the unit, only two, understanding and using major directions in description of relative position, and observing and identifying motion relative to Mr. O were attained successfully. Group tests on relative position were administered to an entire class of children at one time. Individual interviews regarding relative motion were carried out with one child at a time in a separate testing room.

³⁹ Abraham S. Fischler, "Change in Classroom Behavior Resulting From an In-Service Program Utilizing Television," School Science and Mathematics, LXVII (April, 1967), 321-24.

⁴⁰ Ellen Siegelman and Elizabeth Karplus, "Relativity Feedback Testing," What Is Curriculum Evaluation?, ed. by Robert Karplus (Berkeley: University of California, 1968), pp. 3-8; Herbert D. Thier, "A Look at First Graders' Understanding of Matter," Journal of Research in Science Teaching, III (March, 1965), 84-89.

Thier investigated first graders' understanding of the concept of matter using the Material Objects unit. He interviewed, individually, 30 first grade students using SCIS materials, and 30 first grade students who were not using SCIS materials. Interviews centered on physical objects such as soda crackers, potassium chloride crystals in water, and sodium bicarbonate reacting with vinegar. Students were asked to describe similarities and differences, explain what happened, and/or describe physical properties. The information gained from these interviews proved helpful in redesigning the Material Objects unit. Thier admitted that the interview tool is time consuming and exhausting, but a promising method for evaluating the effects of the SCIS program. Also, it provides insight into the preconceptions held by these children about those natural phenomena which are not a part of the teaching program but which are presented during the interview.

Battaglini undertook a study to determine fourth graders' ability to understand concepts of relative position and motion by using the planetarium as a testing device.⁴¹ With the planetarium-oriented evaluation process, the student is shown examples of relative position and motion that are different from those examples previously seen in the classroom.

⁴¹Dennis W. Battaglini, "An Experimental Study of the Science Curriculum Improvement Study Involving Fourth Graders' Ability to Understand Concepts of Relative Position and Motion Using the Planetarium as a Testing Device" (unpublished Ph.D. dissertation, Michigan State University, 1971).

A 30-item planetarium test was created and used as a pre- and post-test, and was administered to nine classes of SCIS fourth grade students and six classes of non-SCIS children. Analysis of covariance was used to test the null hypothesis of no significant difference in adjusted mean scores between the SCIS and non-SCIS groups.

Both groups of students improved their mean score on the post-test. However, the SCIS students showed a gain from pre-test to post-test that was more than twice as large as the non-SCIS group gain. The analysis indicates that this difference is significant at the 0.05 level; thus the null hypothesis was rejected. It was concluded that the fourth graders enrolled in the SCIS program after having received the material presented in the unit titled Relativity had a significantly greater ability to understand the concepts of relative position and motion than a comparable group of students who had not received such instruction.

Allen compared first grade SCIS students with first grade non-SCIS students in their ability to describe objects by their properties.⁴² Nineteen property categories, five of which are covered in the SCIS unit Material Objects, were offered by the children who were tested. Each child was tested individually in a "clinical" situation. The subject was given a soda cracker and asked to state as many properties

⁴²Leslie R. Allen, "An Examination of the Ability of First Graders From the Science Curriculum Improvement Study Program to Describe an Object by its Properties," Science Education, LV, 1 (1971), 61-67.

of this object as possible. There was no time limit. Of the five properties mentioned in the unit, he found that three of these properties involved significantly more SCIS than non-SCIS children. SCIS children appear to be able to apply to a new situation specific property words learned in the classroom.

It is interesting to note that a search of the literature did not reveal any attempts to evaluate cumulative aspects of the SCIS program, i.e., acquisition of objectives over a period of four or five years. Further, no standardized, paper and pencil instrument has been developed to measure the degree to which students have met some of the objectives of the SCIS program. One possible reason for this has already been stated: It is the philosophy of the SCIS program that comparisons among children seem completely irrelevant. Also, a child should be evaluated in terms of his individual progress in the attainment of specified goals, not in terms of how he compares with another student in reaching these goals.⁴³ Experimental studies involving SCIS thus far appear to respect this philosophy.

Development of Instruments to Measure Process Skills

Some early attempts at constructing instruments to measure problem-solving ability include the Balanced Problem

⁴³ Science Curriculum Improvement Study, SCIS Newsletter, p. 3.

Test developed by Cross and Gaier.⁴⁴ The test consists of six sets of problems, each set containing four problems. Each set of problems may be solved either through applications of facts for each specific problem or through the use of a general principle applicable to an entire set of problems. Both principle and facts on a given page are covered by irreplaceable tabs. While the development of the Balanced Problem Test concerns itself with problem solving in mathematics at the high school level and thus does not have any specific relation to this study, it does represent an attempt to develop a problem-solving test.

The Questest developed by Suchman utilizes special films to present problem episodes to individual elementary school children.⁴⁵ It consists of (1) a problem episode, (2) a period of instruction, (3) a period of inquiry, (4) a set of follow-up paper and pencil tests to measure information and conceptual growth, and (4) a coding and scoring system for the analysis of the structure and content of the questions asked. The test is given to one child at a time under carefully controlled conditions. The data obtained from the tape-recorded protocols of the inquiry periods present almost infinite possibilities for analysis. For the

⁴⁴K. Patricia Cross and Eugene L. Gaier, "Technique in Problem Solving as a Predictor of Educational Achievement," Journal of Educational Psychology, XLVI, 4 (1955), 193-206.

⁴⁵J. Richard Suchman, The Elementary School Training Program in Scientific Inquiry (Urbana: University of Illinois, June, 1962).

purposes of his study, he limited the analysis to the strategies of question design and application in data gathering.

More recent efforts in developing instruments to measure process or problem-solving skills include those developed by the American Association for the Advancement of Science (AAAS), which has developed an evaluation model for the elementary program Science--A Process Approach, on the behavioral position.

Unlike those who favor evaluation of content, the curriculum organization which appeals to a behavioral interpretation approaches the development of instructional materials by asking one question: "What do we want the learner to be able to do after instruction that he was unable to do before instruction?"⁴⁶

Competency measures, consisting of tasks intended to assess the achievement of the objectives for each exercise, have thus been developed. The competency measures for the first four parts of Science--A Process Approach are for the assessment of individual children. For parts 5, 6, and 7, the individual form of the test is turned into a form for group administration, and both forms are provided.⁴⁷

While the group competency measures for grades 4, 5, and 6 have been so designed that any number of students can be tested simultaneously, administration of these

⁴⁶Henry H. Walbesser, "Science Curriculum Evaluation: Observations on a Position," The Science Teacher, XXXIII (February, 1966), 34.

⁴⁷American Association for the Advancement of Science, Science--A Process Approach: An Evaluation Model and Its Application, Second report (Washington, D. C.: American Association for the Advancement of Science, Misc. Pub. 68-4, 1968), p. 13.

competency measures may be somewhat cumbersome for the classroom teacher. The teacher has to read each question aloud while the members of the class follow along on their sheets. Time is allowed for each student to answer the question before the teacher reads the next question. Some of the exercises require materials other than paper and pencil for each student. The answer sheets are not machine scorable because many answers, especially those that require the student to make a diagram or picture, require interpretations made by the teacher.

Beard suggested the possibility that the competency measures for parts 1-4 of Science--A Process Approach would not be used regularly because they are administerable only individually.⁴⁸ Thus, she undertook the task of writing a group process test for grades 1-3 for Science--A Process Approach. A multimedia test format was used to evaluate the attainment of two processes: measuring and classifying. Pupils viewed and listened to each item. This format attempted to assess science process skills of groups of primary (grades 1, 2, and 3) pupils who had not yet perfected their reading and writing skills sufficiently to use a written test. A series of 35-millimeter colored slides illustrating laboratory situations involving basic processes was presented, synchronized with a tape recording which orally provided instructions and stated the problem to be

⁴⁸Beard, "The Development of Group Achievement Tests," pp. 179-83.

considered. Each pupil indicated his answer to each question by marking his answer sheet as directed by the tape recording. Thus, all pupils had an equal opportunity to consider each item.

The chronology of developmental procedures used by Beard in her study was as follows:

- (1) Selection of representative basic science processes
- (2) Design of the answer sheet format
- (3) Design and construction of test items
- (4) Validation of test items
- (5) Development of general instructions and sample items
- (6) Pilot testing with students of appropriate minimum ages
- (7) Pilot testing with students who had experience in S--APA and who were of appropriate minimum ages
- (8) Assembly of test items into sample BSPTs
- (9) Production of synchronized tape recordings
- (10) Administration of sample BSPTs
- (11) Scoring of BSPTs and analysis of their performances
- (12) BSPT item analysis from test data only
- (13) Analysis of BSPT item group performances⁴⁹

Evidence from Beard's study indicates that a multimedia test format can be made to function effectively as a group test for primary grade pupils. Two of the six sample BSPTs were reliable as indicated by test-retest correlation of 0.70 or greater. These two tests also discriminated between students with and without the Science--A Process Approach experience, supporting the claim of content validity. The validity of the test remains in doubt, however, as the items were not evaluated against the actual behaviors of the children.

⁴⁹Ibid., p. 180.

While she suggested that group tests may not be as sensitive or thorough as individual performance tasks, she suggested that much process achievement information can be gathered efficiently with a properly constructed group test.

Fyffe and Robison developed test items for the integrated processes of "formulating hypotheses" and "defining operationally."⁵⁰ Multiple choice items were developed using the behavioral objectives of Science--A Process Approach as a basis, into a 79-item group test.

Individual testing of 56 students using individual competency measures from Science--A Process Approach was completed, then followed by administration of the group test to those same students. The scores on the individual competency measures served as the external criterion measure for selection of the upper and lower 27 per cent categories necessary to calculate group test item discrimination indices.

Items with indices of item discrimination of 0.20 or greater were then considered as a sub-test of the 36 items for each of the two processes under consideration. Student answer sheets were scored using only those items. From the two sets of scores, individual competency measures and sub-tests from the group test items, Pearson product-moment

⁵⁰ Darrel W. Fyffe, "The Development of Test Items for the Integrated Science Processes: Formulating Hypotheses and Defining Operationally" (unpublished Ph.D. dissertation, Michigan State University, 1971); Richard Wayne Robison, "The Development of Items Which Assess the Processes of Controlling Variables and Interpreting Data" (unpublished Ph.D. dissertation in process, Michigan State University, 1970).

correlation coefficients were computed. It was found that individual competency measure scores for the two integrated processes are correlated, at the .001 level of significance, with their representative items from the group test.

A study of the results leads to the conclusion that it is possible to measure the acquisition of skills in the integrated science processes, Formulating Hypotheses and Defining Operationally, if selection of test items is based upon reference to a criterion measure.

A study by Walbesser and Carter deals further with this question of group versus individual competency measures.⁵¹ At the time of their study, editions of AAAS Science--A Process Approach included competency measures for grades K-3, the first four parts in the instructional sequence. As previously mentioned, these measures are individually administered. During the field testing of this program, the teachers requested a more convenient competency measure be developed for parts E and F (grades 4 and 5) of the Science--A Process Approach instructional sequence. Thus, individual and group measures were constructed.

Nine performance classes were named: identifying, ordering, constructing, describing, demonstrating, stating a rule, applying a rule, and distinguishing. Behavioral

⁵¹Henry H. Walbesser and Heather L. Carter, "The Effect on Test Results of Changes in Task and Response Format Required by Altering the Test Administration From an Individual to a Group Form," Journal of Research in Science Teaching, VII, 1 (1970), 1-8.

objectives for each of these classes were formulated for both individual and group forms.

On the basis of the perceived differences between the individual and group competency tasks, the following hypothesis was made: The difference in performance results for a given objective within a given process is as likely to favor the group competency measure data as it does the individual competency measure data.⁵²

To test this hypothesis, group and individual competency measure data were gathered during the period 1966-68 from random samples of students in each of the 14 Science--A Process Approach test centers located around the United States. The tryout teachers for each part were randomly assigned to one of two test administration schedules. One schedule began with the group form for the first exercise competency measure, the individual form for the second exercise, and continued to alternate group and individual for all exercises taught. The other schedule was the opposite, with the individual form for the first exercise.

It appears from the data that the individual competency measure results are consistently higher than the group competency measure results. The authors further suggested that while direct performance-based assessment does take more time, it may be a worthy investment.

⁵²Ibid., p. 4.

Dietz and George attempted to develop a reliable and valid group test that would evaluate some of the problem-solving skills of children in grades 1, 2, and 3.⁵³ These skills were identified as: (1) the ability to recognize the problem presented; (2) an understanding of science principles needed to solve the problem; (3) the ability to collect data; and (4) the ability to reason with "if-then" statements.

Both forms of the Problem Solving Skills Test contain two picture-problems dealing with physical science principles. Each problem is presented to the class in the form of a large picture. To minimize variance due to reading, the directions and the test questions are posed orally and each student is supplied with answer sheets containing only pictures. The child indicates his answer by placing an "X" in a box beneath one of the picture choices given for each question.

The test is not curriculum specific so that it can be used to evaluate students' problem-solving abilities regardless of what particular science programs the student has studied in school. Generally, the data compiled using the PSSST seem to indicate that the test is ". . . evaluating what it was intended to evaluate."⁵⁴ Limitations of the test include the fact that the norms must be used separately according to area, i.e., suburban, outlying urban, and urban.

⁵³Dietz and George, "A Test to Measure Problem Solving Skills," pp. 341-51.

⁵⁴Ibid., p. 350.

Also, there is a lack of external criteria for determining the validity of the test items and the validity of the identified skills as a part of the problem-solving process.

The Test of Science Processes developed by Tannenbaum⁵⁵ was designed to assess achievement and diagnose weaknesses in the ability of students in grades 7, 8, and 9 to use the processes of observing, comparing, classifying, quantifying, measuring, experimenting, inferring, and predicting.

The instrument has the following characteristics:

- (1) It consists of 96 multiple choice (five choice) questions;
- (2) It requires total actual testing time of 73 minutes (some students may finish in less); (3) the test booklet is printed with black and white illustrations and, for the 12 questions which require color, 35-millimeter color slides are used;
- and (4) Scoring of the instrument yields a total score (Kuder-Richardson Formula #20 reliability = 0.91) and eight subscores, one for each process (reliabilities from about 0.30 to 0.80).⁵⁶

The procedures that Tannenbaum used in constructing his test are summarized as follows:

- (1) Definition of and statement in behavioral terms of those science processes which it was appropriate to expect 7th, 8th and 9th graders to be able to use in the light of current science education.

⁵⁵Robert S. Tannenbaum, "The Development of the Test of Science Processes," Journal of Research in Science Teaching, VIII, 2 (1971), 123-36.

⁵⁶Ibid., p. 132.

- (2) Evidence of the curricular validity of the statement of science processes was gathered by submitting it to experts in science education.
- (3) Form I, consisting of 98 items, was written. Each item consisted of a 35 mm color slide visual stimulus, a mimeographed stem, and five choices. Simultaneous vocal stimulus gave instructions and read the questions to minimize reading problems.
- (4) Administration of Form I to 156 students in grades 7, 8, and 9 from a Bronx, New York, public intermediate school.
- (5) Results of Form I subjected to item analysis.
- (6) Form II written, consisted of 96 multiple choice items. The 96 items and printed instructions were typewritten, and the color slides (except for questions 1-12 on Form II) were converted to tables or black and white prints.
- (7) Administration of Form II to 3,673 students.
- (8) Results of Form II used to determine the norms, item statistics, reliabilities, and validities.
- (9) Construction of test manual and inclusion of statistical tables.⁵⁷

There is considerable evidence of both the content and the curricular validity of the test, due to the care employed in the construction and validation of the blueprint and items. Predictive validity has not yet been investigated. Criterion-related validity was attempted using a small population ($N = 35$) who took the TSP, and their teacher, who had observed their classroom and laboratory behavior for more than a full semester, and rated them individually on a scale of zero to nine on each of the eight processes. While the evidence from this investigation was not unequivocal, it may be interpreted as at least some indication of a degree of criterion-related validity where the criterion is teacher rating of the students' ability to use the processes.

⁵⁷Ibid., pp. 124-25.

The very existence of an instrument of this nature (i.e., objective testing of process-related behavioral objectives) should demonstrate to science educators the feasibility of the development of other such instruments and of objective studies in this area.

The TAB Science Test, developed by Butts and Jones, represents a unique way of evaluating problem-solving behavior.⁵⁸ It is designed to analyze the methods as well as the end products of inquiry by investigating the type, number, and sequence of questions asked by elementary school children when solving problems.

The test uses the tab-item format in which the student is given a problem, and list of clue questions which may be used to help him solve the problem. The answer to each question is covered with a tab. If the student wishes to ask a question, he is free to remove the corresponding tab to find the answer. The specific tab and the order in which tabs are pulled are recorded on the student's answer sheet. With this format, the type and sequence of tabs pulled give an indication of the problem-solving or inquiry behaviors of the subject.

Two forms of the TAB Science Test were developed, and tested with 2,519 fourth, fifth, and sixth grade students in six Texas Independent School Districts. In each of these school districts a wide range of socioeconomic backgrounds,

⁵⁸Butts and Jones, "The Development of the TAB Science Test," pp. 463-73.

tested intelligence, science knowledge, and reading scores were exhibited by the subjects. Scoring of the tests took into consideration the number, type, and value of clues obtained by the child. Also considered was the order of the tab pull.

The validity of the TAB Science Test was determined by using a Chi-square analysis of the relationships between the TAB Science Test scores and student behaviors of searching, data processing, verifying, discovering, assimilating, and accommodating, which were predicted from the model of inquiry. In addition, using a Rank Correlation analysis, a relationship between TAB Science Test scores and teacher's rating of students was found.

The reliability of the TAB Science Test was determined by calculation of coefficients of equivalents (0.365 and 0.420) and coefficients of internal consistency (0.497 and 0.532).⁵⁹

The TAB Science Test appears to be a meaningful tool by which to appraise student behaviors. It is one way to define, isolate, and measure the processes by which a problem is solved, rather than securing just the solution to the problem.⁶⁰

⁵⁹David P. Butts, An Inventory of Science Methods (Austin, Texas: Science Education Center, University of Texas, 1966), p. 25.

⁶⁰Ibid.

A Test of Volume Concepts and The Learning Hierarchies Test were developed in a study by Howe and Butts to study the effects of instruction based on Science--A Process Approach on children's learning of selected concepts of volume.⁶¹ In both tests, the examiner demonstrates problems with objects to groups of children; the children mark individual response sheets. Fourth and sixth grade children with approximately two years of experience in Science--A Process Approach and a similar group of students with no experience in this program were pre-tested on their knowledge of displacement of volume and volume as occupied space, as well as with the Learning Hierarchies Test, instructed in concepts related to volume, and post-tested using the same instruments.

Results showed that fourth grade children who experienced instruction based on Science--A Process Approach scored higher on the volume tasks in the pre-test than children who had not experienced it. All groups experienced in Science--A Process Approach outscored the inexperienced groups on the Learning Hierarchies Test during the pre-test. No other significant differences were found.

Edwards developed 18 lessons for kindergarten children in four intellectual skills: classification, measurement,

⁶¹Ann C. Howe and David P. Butts, "The Effect of Instruction on the Acquisition of Conservation of Volume," Journal of Research in Science Teaching, VII, 4 (1970), 371-75.

observation, and data treatment.⁶² He then developed group nonreading tests which would measure the growth of these four intellectual skills. His Skills Test, which he developed for the study, was used as a pre-test and a post-test for students who were given the lesson.

The format of the test is such that oral questions are posed by the examiner, and the student responds to the question by writing or coloring in the corresponding page of the answer booklet. The pages in the booklet are of three different consecutive lengths, and are in three different colors, to help the students and the tester to keep together.

A control group also took the test at the time of the experimental group post-testing. He concluded that the materials and the test used were appropriate for kindergarten children. He also suggested that the lessons studied have a positive effect on reading readiness scores of participating youngsters.

Morgan utilized eight-millimeter motion picture film loops as the medium for developing a diagnostic test for sixth, seventh, and eighth grade pupils.⁶³ The STEPS (Science Test for Evaluation of Process Skills) test was

⁶²Thomas F. Edwards, "A Study in Teaching Selected Intellectual Skills to Kindergarten Children" (unpublished Ed.D. thesis, Michigan State University, 1966).

⁶³David A. Morgan, "STEPS--A Science Test for Evaluation of Process Skills," The Science Teacher, XXXVIII (November, 1971), 77-79.

designed in his study to evaluate competency in the science processes based on student interaction with real materials and equipment. Film loops depicting experiments involving pendulums, magnets, thermal expansion, Archimedes' principle, and Boyle's law were shown to groups of 50 students. Multiple choice questions relating to the films were then answered by the students in the test booklet.

The differences between the means for the sixth and seventh graders were found to be significant at the 0.05 level or better. Differences between means for seventh and eighth graders were significant at the 0.05 level for all processes.

Methods Used in Test Construction

Many references are made in the literature to the necessary steps one must take in constructing a test. These served as guidelines in the development of the Test of Science Inquiry Skills.

Preliminary Considerations

The first step in construction of a process-oriented test is to list and define the processes to be evaluated.⁶⁴ This applies to all tests, whether they be concept or fact

⁶⁴Marjorie W. Geesaman, "Evaluating the Educational Outcomes of Process Education in Science by Objective Test" (paper presented at the 20th meeting of the National Science Teachers Association, New York, April, 1972), p. 2; Leo Nedelsky, Science Teaching and Testing (New York: Harcourt, Brace & World, Inc., 1965), p. 149; Norman E. Gronlund, Measurement and Evaluation in Teaching (2nd ed; New York: The MacMillan Company, 1971), p. 129.

oriented. The first requisite for writing a test for a course is a clear description of the subject matter and abilities that are the objectives of the course.

Many authors are of the opinion that these processes should be stated behaviorally; they should be identified and defined in terms of desired changes in pupil behavior.⁶⁵ Evaluation should assess the success of instruction in terms of measuring the behaviors acquired by each child as set forth in the objectives.⁶⁶ Not only do behavioral objectives give direction to test construction, but also to all phases of course planning and teaching.⁶⁷

The next procedure in construction of a process-oriented test is to describe the content within which the skills are to be tested.⁶⁸

Once the first two steps have been accomplished, a table of specifications can be drawn up.⁶⁹ If the processes

⁶⁵Gronlund, Measurement and Evaluation, p. 129.

⁶⁶Walbesser, "Science Curriculum Evaluation," p. 38.

⁶⁷Clarence H. Nelson, Measurement and Evaluation in the Classroom (Toronto: Macmillan Company, 1970), p. 21.

⁶⁸Geesaman, "Evaluating Educational Outcomes," p. 2; Gronlund, Measurement and Evaluation, p. 129; Benjamin S. Bloom, J. Thomas Hastings, and George F. Madaus, Handbook on Formative and Summative Evaluation of Student Learning (New York: McGraw-Hill, 1971), p. 28.

⁶⁹Nelson, Measurement and Evaluation, p. 50; Gronlund, Measurement and Evaluation, p. 129; Geesaman, "Evaluating Educational Outcomes," p. 2; Bloom, et al., Handbook on Evaluation, p. 28; Nelson, personal communication; Benjamin S. Bloom, et al., Taxonomy of Educational Objectives Handbook I: Cognitive Domain (New York: David McKay Company, Inc., 1956).

are listed down one axis of the chart and the content categories are listed across the other axis, a two-dimensional test specification grid is produced. By inserting numbers of items to be written at the intersection of each process and content category, one can design an entire test. The relative emphasis that each content category and each objective should receive is indicated by the number of items allotted to the various cells on the grid.

Without a carefully developed test plan, ease of construction all too frequently becomes the dominant criterion in selecting and constructing test items. As a consequence, the test measures a limited and biased sample of pupil behavior and neglects many of the learning outcomes considered most important by the teacher. In short, without a carefully developed test plan, the test tends to lack content validity.⁷⁰

Item Writing

Nelson listed the following general considerations in the writing of test items:

- (1) Every item in the test should measure, in an incisive manner, something of consequence or significance in the course or area.
- (2) Every item should have an indisputable right answer.
- (3) Every item should be stated in clear, concise, precise, straightforward and grammatically accurate phraseology.
- (4) Every item should be free of any irrelevant clues to the answer, and no item should reveal the answer to any other item in the test.

⁷⁰Gronlund, Measurement and Evaluation, p. 130.

- (5) Every item should be independent. One item in a test should not depend on another item for its meaning.⁷¹

Geesaman pointed out further that the special aspect of items which evaluate processes is that they should provide all the data the student needs to perform the process named.⁷² They do not require that the student bring information to the test. It was found that an economical method of testing many processes and content areas in one test is to present one stimulus, such as a map, graph, chart, diagram, or description of an experiment, and design several items related to that stimulus. Each of these items could test a different process. The next stimulus could deal with a different content area and test the same or a different assortment of processes.

Gronlund termed this the "interpretive exercise," in which pupils are required to demonstrate the specific interpretive skill being measured.⁷³ For example, pupils might be called upon to recognize assumptions, inferences, conclusions, relationships, applications, and the like.

Multiple Choice Items

The multiple choice type item is one of the most widely used in test construction. It consists of an introductory statement or question, usually referred to as the

⁷¹Nelson, Measurement and Evaluation, pp. 53-54.

⁷²Geesaman, "Evaluating Educational Outcomes," pp. 2-3.

⁷³Gronlund, Measurement and Evaluation, p. 216.

stem, followed by three, four, or five responses. All the responses except the correct answer are referred to as foils or distractors. Many teachers limit the use of multiple choice items to the knowledge area because they believe that all objective type items are restricted to the measurement of relatively simple learning outcomes. Gronlund pointed out that multiple choice type items may be effectively adapted to measure the ability to: (1) apply facts and principles, (2) interpret cause-and-effect relationships, and (3) justify methods and procedures.⁷⁴

The Biological Sciences Curriculum Study (BSCS), in constructing a resource book of test items, chose the multiple choice format because it presents a versatile means of gathering data on student performance.⁷⁵ The format is versatile because questions can be formulated on nearly any subject and the items can be designed to measure various levels of cognitive ability.

In comparing multiple choice items with other items such as the true-false or fill-in type item, Nedelsky stated that

Since the relative correctness and attractiveness of the responses clearly depend on the question asked, the multiple-choice form affords better control over them and thus over the mental processes of the student.

⁷⁴Ibid., pp. 179-83.

⁷⁵Biological Sciences Curriculum Study, Resource Book of Test Items for Biological Science: Molecules to Man (Rev. ed.; Educational Programs Improvement Corporation, 1971), p. xiii.

It is therefore the most flexible and useful form of objective exercises and allows a more precise definition of the objective being tested.⁷⁶

Ebel described multiple choice items as being adaptable to the measurement of most important educational outcomes--knowledge, understanding, and judgment; ability to solve problems, to recommend appropriate action, to make predictions.⁷⁷ Almost any understanding or ability that can be tested by means of any other item form--short answer, completion, true-false, matching, or essay--can also be tested by means of multiple choice test items.

General guidelines for writing multiple choice items, as outlined by Nelson, include the following:

- (1) The stem should present a complete premise and therefore must contain a verb.
- (2) The responses should be parallel in construction and length.
- (3) The correct answer should not be conspicuously longer or more elaborate than the other responses.
- (4) Care should be taken to avoid giveaway clues to the answer.
- (5) Excess verbiage should be eliminated, retaining only the minimum required for the student to answer the item.
- (6) Each multiple choice item should have one, and only one, scorable right answer, though multiple elements can be incorporated into a single answer.⁷⁸

⁷⁶ Nedelsky, Science Teaching and Testing, p. 157.

⁷⁷ Robert L. Ebel, Essentials of Educational Measurement (2nd ed.; Englewood Cliffs, New Jersey: Prentice Hall, 1972), p. 187.

⁷⁸ Nelson, Measurement and Evaluation, p. 75.

Experimental Tryout of Test Materials

Very few specific guidelines for pilot testing of an instrument such as the one developed in this study were found in the literature. While the chronology involved in the tryout of a test from other similar studies mentioned in this chapter proved helpful (examples: Beard, Tannenbaum), a description by Conrad represents the most concise set of guidelines found.⁷⁹ He subdivided the process of experimental tryout into three stages: the pretryout, the tryout proper, and the final administration.

(1) Pretryout: This is the preliminary administration of the tentative tryout units to small samples of examinees for the purpose of discovering gross deficiencies, but with no intention of analyzing pretryout data for individual items. It may be highly informal, and may involve examinees fairly representative of the population to whom the finished test is to be administered. Major omissions, ambiguities, or inadequacies in the directions to the examinees and in the simple items and fore-exercises may be discovered in the pretryout.

(2) Tyrount: Once the gross deficiencies in the tryout forms have been eliminated, it becomes necessary to obtain accurate information concerning the performance of each item in a sample of examinees similar to those with

⁷⁹Herbert S. Conrad, "The Experimental Tryout of Test Materials," Educational Measurement (Washington, D. C.: American Council on Education, 1951), pp. 250-65.

whom the final form of the test is to be used. Administration of the tryout forms of the test for this purpose to 400-500 or more examinees is regarded as constituting the tryout proper. In some cases, a second tryout is necessary to ascertain the adequacy of items revised after the first tryout.

(3) Trial Administration of the Finished Test: On the basis of the data obtained in the tryout, the items are selected and assembled into the finished test. This trial administration serves to indicate exactly how the test will function in actual use. This means that no material changes can be made after the trial administration, and that the sample employed must be essentially like the group with whom the test is to be used.

In addition to the above guidelines, Conrad gave further suggestions on the following:

(1) Sampling: The sample of examinees must be essentially similar to those with whom the test is to be used, in order to avoid bias in the data.

(2) Administrative arrangements: The tryout materials should be administered under as nearly as possible the same conditions as those under which the finished test will be administered. This includes consideration of: (a) distribution of tryout materials to participating schools, (b) examiners, (c) directions to examiners, and (d) directions to examinees.

Results of the trial administration should include besides the test scores, feedback from schools, teachers, and students involved.

Reliability, Validity, Item
Difficulty and Discrimination

Appropriate indices which were used in this study and methods of computing them, will be discussed in Chapter IV.

Studies Which Compare Problem Solving With
Achievement and Attitude

Some of the hypotheses tested in this study involved the correlation of results from the Test of Science Inquiry Skills with performance on other instruments such as the Science and Scientists Attitude Inventory, the STEP test, and the State of Michigan Assessment composite achievement score.

Other studies have also attempted to determine correlations among these aspects. Butts investigated the relationship between (1) knowledge of science facts and principles (achievement) and (2) problem-solving behavior among college seniors.⁸⁰ The STEP test was utilized to secure data regarding knowledge of science facts and principles. The X-35 Test of Problem Solving, an instrument using the tab-item format, was developed by Butts for the study to assess

⁸⁰ David P. Butts, "The Relationship of Problem-Solving Ability and Science Knowledge," Science Education, XLIX, 2 (1965), 138-46.

behavior in problem solving. The findings reflected a mean coefficient between the scores on the STEP and the X-35 Test of 0.14. The error variance was ± 5.59 . Statistically, the findings show that the instrument may have little or no value because of a large error variance.

Welch and Pella, in developing the Science Process Inventory (SPI) for secondary school students, obtained correlation coefficients between SPI scores and scores obtained on a test of general mental ability (Henmon-Nelson Test of Mental Ability) and reading ability (STEP reading test and Iowa Test of Educational Development reading test).⁸¹ Coefficients of 0.61, 0.64, and 0.62 for three grade levels were calculated between SPI and the Henmon-Nelson Test scores. Coefficients between the SPI and the STEP and ITED tests were 0.65 and 0.66, respectively.

Raun and Butts investigated the hypothesis that if exposed to situations which focus on inquiry and student involvement, then certain changes should occur in student cognitive and affective behaviors as a consequence of interacting with the strategies of inquiry of a curriculum.⁸²

⁸¹Wayne W. Welch and Milton O. Pella, "The Development of an Instrument for Inventorying Knowledge of the Processes of Science," Journal of Research in Science Teaching, V, 1 (1967-68), 64-68.

⁸²Chester E. Raun and David P. Butts, "The Relationship Between the Strategies of Inquiry in Science and Student Cognitive and Affective Behavioral Change," Journal of Research in Science Teaching, V, 3 (1967-68), 261-68.

Pre-test IQ data on the 95 subjects were taken from their school records. To assess specific behaviors of the students, 14 instruments including the California Mental Maturity Test (three forms), Metropolitan Achievement Test (three forms), Semantic Differential (attitude inventory), STEP (two forms), TAB Science Test, Guilford Factor Test (three forms), and the AAAS student Competency Measures were used. The evidence indicates that performance in selected strategies of inquiry is correlated with those behavior factors associated with intelligence, divergent thinking, attending, science recall, reading, and attitudinal perception of the "potency of science." The strategy of inquiry which indicated the greatest number of correlations with behavioral change was "using numbers," followed by "classifying," "space-time relations," and "observing."

A study by Klopfer attempted to evaluate the effectiveness and effects of Charting the Universe, Book 1, prepared by the University of Illinois Elementary School Science Project (ESSP).⁸³ It included the following hypothesis: "Subject matter achievement is positively correlated with a student's general scholastic ability." Pre- and post-test data from specifically constructed subject matter achievement tests and IQ scores from the Henmon-Nelson Test

⁸³ Leopold E. Klopfer, "Effectiveness and Effects of ESSP Astronomy Materials--An Illustrative Study of Evaluation in a Curriculum Development Project," Journal of Research in Science Teaching, VI, 1 (1969), 64-75.

were analyzed to test this and other hypotheses. Three partial correlation coefficients were obtained for the Book 1 items, for the general knowledge items, and for a total score which combined the first two: 0.260, 0.294, and 0.308. All were significant. Both on items relating particularly to the ESSP materials and on items measuring general knowledge about astronomy, achievement was related positively, though only slightly, to general scholastic ability.

Butts and Jones investigated the relationship between inquiry training and concept development in sixth grade students.⁸⁴ Pre-tests included the following battery of tests: STEP, National Achievement Test (Elementary Science Test), and the TAB Inventory of Science Processes (TISP). The TISP, developed specifically for the study, is based on the tab-item format, and is designed to assess the steps by which a student solves a problem. It appears that there is no relationship between inquiry training and changes in students' recall of science factual knowledge as measured by these instruments. This may have been a function of the selectivity and size ($N = 109$) of the sample. Further study needs to be done with a larger group of students in a less homogeneously selected situation.

⁸⁴David P. Butts and Howard L. Jones, "Inquiry Training and Problem Solving in Elementary School Children," Journal of Research in Science Teaching, IV, 1 (1966), 21-27.

In another study by Butts and Jones, the STEP was utilized as an external criterion.⁸⁵ A validity coefficient of approximately 0.05 between the TAB test designed specifically for the study, and the science reasoning questions of the STEP was calculated. This indicated that the STEP may have been inadequate as an external criterion because the correlation was so low.

Taylor outlined an analysis for efficient evaluation of science instructional programs.⁸⁶ Evaluation must include dimensions other than cognitive; particularly those which affect changes in behavior, such as social and physical. Whether or not desired social changes, or changes in attitudes have occurred is a matter to be decided by the teacher, and such decisions must be supported by records. Changes in student comprehension of data in science on the other hand can be determined by comparison with local, state, and national norms. Another method of evaluation is to compare student achievement before and after the planned learning experiences.

In the realm of attitudes of students toward science and scientists, the literature is lacking in studies which specifically compare student attitude with achievement and/or problem-solving ability. An early attempt prior to the

⁸⁵Butts and Jones, "The Development of the TAB Science Test," pp. 463-73.

⁸⁶Alton L. Taylor, "Curriculum and Instructional Evaluation in Science," Science Education, LIV, 3 (1970), 237-39.

"Sputnik Era" to measure secondary students' attitudes toward science and scientists was carried out by Mead and Métraux under the sponsorship of the American Association for the Advancement of Science.⁸⁷ Students were asked to write a brief essay on a topic set by an incomplete sentence which was printed at the top of the page. The total sample of essays (1,000) was drawn randomly from approximately 35,000 essays. While the image of the scientist was shown to be a positive one held by the students, a negative attitude as far as his personal career or marriage is concerned, was revealed by this study.

Studies by Motz and Glass involved the development of attitude inventories, but these inventories were not correlated with existing achievement or problem-solving tests.⁸⁸

Stoker surveyed the aptitude and attitudes of high school youth toward science and scientists and the relationships of these factors to each other and to the personal traits of youth.⁸⁹ Among 2,500 pupils surveyed in grades 10

⁸⁷ Margaret Mead and Rhoda Métraux, "Image of the Scientist Among High School Students," Science, CXXVI (August, 1957), 384-90.

⁸⁸ LaMoine L. Motz, "The Development of an Instrument to Evaluate Sixth and Ninth Grade Students' Attitudes Toward Science and Scientists" (unpublished Ph.D. thesis, University of Michigan, 1970); Lynn W. Glass, "Assessment of Affective Outcomes of Instruction With High School Sophomore Biology Students and Teachers," Journal of Research in Science Teaching, VII, 1 (1971), 41-43.

⁸⁹ Howard W. Stoker, "Aptitudes and Attitudes of High School Youth in Regard to Science as Related to N Variables" (unpublished Ph.D. dissertation, Purdue University, 1957).

through 12, a favorable attitude toward science as a social institution was generally expressed. A significant relationship was reported between aptitude and attitude toward scientists as people. Attitudes toward science as an institution and as a vocation and attitudes toward scientists were closely related to students' grades in science and their socioeconomic status.

Lyda and Morse noted not only changes in attitudes toward arithmetic when meaningful methods of teaching were used, but also significant gains in arithmetic computation and reasoning associated with methods and attitudes.⁹⁰

While these two studies represent attempts at correlating attitude with achievement, one involves the area of mathematics education and the other involves high school science; therefore, neither has any direct bearing on this study.

More recently, Partin compared achievement and interest in science between fourth graders taught Science--A Process Approach and a control group of fourth graders who used a textbook.⁹¹ Criterion measures were the California Achievement Test, the Sequential Test of Educational

⁹⁰ Wesley J. Lyda and Evelyn C. Morse, "Attitudes, Teaching Methods, and Arithmetic Achievement," Arithmetic Teacher, VIII (March, 1961), 117-19.

⁹¹ Melba S. Partin, "An Investigation of the Effectiveness of the AAAS Process Method Upon the Achievement and Interest in Science for Selected Fourth Grade Students" (unpublished Ph.D. dissertation, University of Southern Mississippi, 1967).

Progress (Science), Competency Measures, and an investigator-developed Informal Interest Inventory. No differences were found between experimental and control groups on the first two tests. However, boys scored higher than girls in these two tests and on the Informal Interest Inventory. The experimental group scored higher on the Competency Measures and the Informal Interest Inventory. Performance on the Competency Measures was unrelated to IQ scores.

Benson compared the cognitive and attitudinal outcomes of two methods of instruction: lecture-demonstration, and pupil-investigatory approaches.⁹² Fifth grade pupils, randomly assigned to experimental and control groups, were taught by the same teachers over a seven-month period. Criterion measures were the Stanford Science Achievement Test and two other tests: Interest and Ideas, and Activities. Results indicated the different methods of presentation produced no significant differences in content achievement. Differences in attitude toward science were significant in one school but not in the second school. Data from the interest inventory indicated no significant difference between sample groups of either school.

It is not uncommon to hear educators speak of the positive correlation between students' attitudes toward science and achievement in science curriculums, but thus

⁹²Keith S. Benson, "A Comparison of Two Methods of Teaching Fifth Grade Science" (unpublished Ph.D. dissertation, Oklahoma State University, 1968).

far, there is little evidence available in the literature that would substantiate this claim.

Summary

A review of the literature concerning elementary science education supports what we see manifested in current elementary science programs--science education has become process oriented. The designing of tests to assess the educational goals being stressed in newer elementary science programs offers a real challenge, especially when instructional emphasis is placed on development of concepts and understanding of the processes of science. Tests should be devised to measure instructional goals not adequately measured by existing instruments.

Reviewing the literature has not only pointed to the need for more appropriate evaluative instruments, but has also helped in the development of the Test of Science Inquiry Skills. Guidelines for planning the test, writing the items, and pilot testing gave valuable direction to this study.

CHAPTER III

DESCRIPTION OF THE STUDY

The purposes of this study were: (1) to develop a fifth grade paper and pencil science process test based on some of the stated or implied goals for grades 3, 4, and 5 of the Science Curriculum Improvement Study; and (2) to use this and other tests to compare fifth grade students who have been in the SCIS program for five consecutive years with fifth grade students who have been enrolled in traditional textbook series science courses.

The General Procedure

The general procedure of developing the Test of Science Inquiry Skills (TSIS) began with the writing of multiple choice test items according to a test specification grid of the process and content areas of grades 3, 4, and 5 of the SCIS program that were to be evaluated. Items were tried out informally on a pre-tryout basis with small samples of students; and more formally during two separate pilot testing programs involving 1,087 fifth grade students who had had varying experiences (traditional and lab-centered) in their science programs. These tryouts provided important feedback information on the format and materials of the test itself. Also, item analyses data from each

tryout provided important information regarding item difficulty and discrimination. Necessary revisions were made after each tryout. The final form of the test reflects this feedback and item analysis data.

The actual study involved the administration of the Test of Science Inquiry Skills to 310 fifth grade students in 12 classrooms who had been in the SCIS program for five consecutive years. A Control group of 195 fifth grade students in seven classrooms was selected on the basis of similarities on composite achievement scores and socioeconomic levels of the 1970-71 State of Michigan Assessment Examination between SCIS and Control schools.

In an attempt to provide a more balanced estimate of students' achievement, the Science Test of the Sequential Test of Educational Progress (STEP) and the Science and Scientists Attitude Inventory (SASAI) were used in addition to the TSIS and the State Assessment composite achievement scores.

Identification of Science Processes

The first step in construction of the process-oriented test was to list and define the processes to be evaluated. In surveying the literature, working with elementary children actively involved in the SCIS program, and in referring to SCIS newsletters and teacher's guides, process and content areas were initially identified as those which are emphasized in the SCIS program in grades 3, 4, and 5.

The SCIS teacher's guides are conveniently divided into four or five major parts, each part containing a number of chapters. The major process and content objectives for each part are clearly stated at the beginning of that part. Further, there is a synopsis of the activities at the beginning of each chapter. Also, by observing children actively involved in the activities of the SCIS program, the processes they were involved in could be identified.

The following list represents the processes of the SCIS program that were initially identified:

- (A) Recognition of the problem
- (B) Formulating a hypothesis
- (C) Identifying and controlling variables
- (D) Interpreting experimental data
- (E) Inferring
- (F) Selecting suitable tests for a hypothesis
- (G) Predicting

Behavioral objectives for each process were then written. Each behavioral objective was designed to indicate what the student must do if and when the particular objective has been met. The following is a list of these behavioral objectives:

(A) Recognition of the problem: After observing the results of interactions of objects and systems of objects in terms of initial and final states, the student can identify a specific problem that can be investigated.

(B) Formulating a hypothesis: After observing a series of interactions, the student constructs a generalization of these observations and states, in testable terms, the primary variable to be investigated.

(C) Identifying and controlling variables: Given a diagram or description of an investigation, the student is able to identify conditions which, when varied, may influence the outcome of an investigation. In a given experimental situation, the student is able to identify the procedure which manipulates only one variable in a systematic way while keeping all other variables that may influence an experimental result constant.

(D) Interpreting experimental data: Given information in the form of tables, charts, graphs, maps, etc., the student is able to analyze this information to formulate inferences, predictions, and hypotheses based on the data presented.

(E) Inferring: By considering the interactions of variables, the student is able to draw warranted generalizations from a body of data. The student is able to determine that variable which is most likely to have caused change in a system.

(F) Selection of suitable tests for a hypothesis: Given a hypothesis to explain a given situation, the student can select a particular empirical approach or series of experiments that can logically verify the hypothesis if it is correct.

(G) Predicting: The student is able to construct a special kind of inference derived from interpolation or extrapolation from a table of data or from a graph.

Identification of Content Areas

The next procedure used in constructing the process-oriented test was to describe the content within which the skills were to be tested. The following content areas are stressed in grades 3, 4, and 5 of the SCIS program:

1.00 Population: When shown several kinds of organisms in a water or land habitat, the student can group like organisms into populations and identify the largest population as the one with the most pictured individuals. The student can also:

- 1.1 Identify predators as animals that eat other animals.
- 1.2 Arrange a series of observable predator-prey relationships into a food chain which accurately depicts the food relationships within the community.
- 1.3 Given a specific community, predict how changing the number of individuals in one population would affect the number of individuals in other populations.

2.00 Subsystems and Variables: The student identifies a group of related objects as a system and can keep track of the original system when the objects are manipulated or added to or removed from the original system.

- 2.1 The student can interpret changes occurring in a system as evidence of interaction.
 - 2.2 The student identifies parts of the original system as subsystems. Given a description or diagram of a mechanical system or liquid system such as a cart, inclined plane, or rubber band plane, the student can:
 - 2.21 isolate subsystems
 - 2.22 identify variables which may affect the system's performance
 - 2.23 design controlled experiments to determine if a given variable does affect a system's performance
 - 2.24 construct or interpret graphs or charts of collected data
 - 2.3 Based on data presented, the student can predict the results of a similar experiment.
- 3.00 Environments:
- 3.1 Environmental factor: Given a diagram or description of an organism living in a particular habitat, the student can identify factors in the environment which may affect the organism's growth, development, and longevity; i.e., water, light, temperature, and chemicals.
 - 3.2 Range of environmental factor: To determine if a particular environmental factor affects an organism, the student describes or

identifies an experimental design which tests one factor over a wide range of intensities.

- 3.3 Optimum range: Given data that indicate a particular environmental factor does affect an organism, the student can identify the portion of the total range where the organism is most successful.

4.00 Relative Position and Motion:

- 4.1 Relative motion: Given a diagram which shows a number of objects in motion, the student can identify those objects which would show motion relative to each other.
- 4.2 Rectangular coordinates: Given a hypothetical map, the student can locate specific places using a rectangular grid.
- 4.3 Polar coordinates: Given a hypothetical map, the student can locate specific places using polar coordinates.

5.00 Communities: When shown several plant and animal populations interacting in a water or land habitat, the student can identify the interdependent populations as a community. Based on information provided through pictures or graphs depicting interactions in a particular community, the student can:

- 5.1 Construct food chains which trace food from plants to plant eaters to animal eaters.

- 5.2 Identify plants as producers and animals that eat plants or other animals as consumers.
- 5.3 Recognize that raw materials contained in dead plants and animals are returned to the community by decomposition caused by bacteria and molds.
- 5.4 Recognize and interpret the cyclic path of food and raw materials in a community.

6.00 Energy Sources: When presented with pictures or descriptions of interactions involving changes in temperature due to friction, force, or sunlight; or speed of objects moving down an incline, the student identifies such interactions as evidence of:

- 6.1 Energy transfer, and can identify the energy source and energy receiver(s) in such a system.
- 6.2 Given a mechanical system such as a cart, inclined plane, or rubber band plane, the student can identify variables which may affect the amount of energy transferred to a given object and can design experiments to manipulate these variables to achieve a desired result.
- 6.3 Based on data from activities with simple mechanical systems, the student can predict the amount of energy transfer that is likely to occur in a given situation.

A two-dimensional test specification grid was then produced by listing the processes down one axis of a chart and the content categories across the other axis. The test was designed by inserting numbers of items to be written at the intersection of each process and content category. Figure 2 depicts the preliminary test specification grid.

Writing of Test Items

Test items were written according to the following guidelines:

(1) One situation would be presented such as a graph, map, chart, diagram, or description of an experiment, and several items relating to that situation would be designed. The special aspect of items which evaluate processes is that they provide all the data the student needs to perform the process named. They do not require that the student bring information to the test.

(2) The situations included on the test would be comparable to, but not identical with, those used in the SCIS program.

(3) The situation provided in the question would offer the student three or four alternatives from which to select the answer.

(4) The total body of items should be designed to match the original process/content test specification grid.

As items were written they were tried out with small samples of fifth grade students mainly for the purpose of

	(A) Recognizing the problem	(B) Formulating a hypothesis	(C) Identifying and controlling variables	(D) Interpreting experimental data	(E) Inferring	(F) Selecting suitable tests for hypotheses	(G) Predicting
1.00 Population 1.1 Predator-prey 1.2 Food chain 1.3 Populations interdependence							
2.00 Subsystems & Variables 2.1 Evidence of interaction 2.2 Identifying systems 2.21 Isolating subsystems 2.22 Identifying variables 2.23 Designing controlled experiments to determine variable's effects 2.24 Constructing and interpreting graphs and charts 2.3 Predicting results of a similar experiment							
3.00 Environments 3.1 Environmental factor 3.2 Range of environmental factors 3.3 Optimum range							
4.00 Relative Position & Motion 4.1 Relative motion 4.2 Rectangular coordinates 4.3 Polar coordinates							
5.00 Communities 5.1 Plant eater-animal eater 5.2 Producers-consumers 5.3 Decomposition 5.4 Cyclic path of materials							
6.00 Energy Sources 6.1 Energy transfer from energy source to energy receiver 6.2 Variables which affect energy transfer 6.3 Predicting amount of energy transfer							

Figure 2.--Preliminary test specification grid.

detecting major problems in readability or clarity. On the basis of this feedback, some of the items originally written were discarded, whereas others were revised and included in the pilot testing program.

Validation of Test Items

A three-member panel of science educators (Dr. Glenn D. Berkheimer, SCIS Trial Center Director at Michigan State University; Dr. Clarence H. Nelson, Evaluation Services, Michigan State University; and Mr. Donald Maxwell, SCIS Implementation Consultant) provided an opinion of validity to each of the items written. If two of the three members agreed that an item was valid, then the item was included on the test. Given the item and the process or content area it represented, the panel graded it as either acceptable, questionable, or unacceptable. Those items judged valid became part of a test item pool of approximately 140 items. These items were then divided into two forms of the preliminary edition of the test of approximately 70 items each.

Instructions and Answer Sheets

Instructions for taking the test were included on the front cover of the preliminary forms A and B and the final form. Before taking the test, the students read these instructions silently while the teacher read them aloud. The instructions also included examples of questions and how the answer sheet should be marked.

The investigator had access to the Evaluation Services at Michigan State University, and therefore MSU machine scorable answer sheets were utilized in the entire study, both pilot and final testing. Students encountered no difficulty in using these answer sheets. Through the State of Michigan Assessment testing program, students had had previous experience with machine scorable answer sheets, and thus feedback from teachers indicated that the students could use these answer sheets successfully.

First Pilot Test

Sample

In October, 1971, the first pilot testing of Forms A and B of the Test of Science Inquiry Skills (TSIS), involving 902 students, was carried out in the school districts of Van Dyke, Livonia, and East Lansing, Michigan, with the following breakdown of students in each district:

	<u>Form A</u>	<u>Form B</u>
Van Dyke	220	214
East Lansing	82	93
Livonia	144	149

Van Dyke elementary programs had, up until the year this study was undertaken, emphasized mainly a textbook-centered science curriculum. Livonia elementary students had been exposed to Science--A Process Approach (AAAS) in grades 1-3, and in grades 4 and 5 to an activity-centered program

developed within the Livonia school district. East Lansing students had been exposed to the SCIS program. Further, those East Lansing teachers involved in the first pilot test had attended a summer SCIS workshop at Michigan State University in the summer of 1971.

Purpose

One main purpose of the initial pilot test was to determine the suitability of the test materials themselves, the instructions for administering the test, and the practicality of the answer sheets.

Also, it was important to obtain item analyses data on each test item from three different populations to determine which items were of suitable difficulty and discrimination levels and those which needed revision or to be eliminated. General criteria were as follows: the item had to have a difficulty below 50 and a discrimination above 40.

Procedure

Forms A and B were arranged in alternating order and compiled into classroom-size packets. Letters to teachers were prepared to explain the purpose of the study and to give instructions. Teachers were encouraged to make suggestions for improving individual items or the entire test.

The preliminary editions of the test were administered by individual classroom teachers. So that data would be available for every item, all the time needed to complete the

test was granted, even if it meant that the students used more than one sitting to complete the test.

All answer sheets were collected and machine scored. Items were then submitted to item analyses.

Results

The instructions provided for test administration were adequate, and the students successfully utilized the machine scorable answer sheets. Many teachers commented that the tests were too difficult for students to read, and took too long to administer. Item analyses data indicated that there were differences in the three populations that took the test. These data also indicated that many items were in need of revision.

It was also suggested that larger type be used on the test. The first preliminary edition of the test used "elite" type. In consideration of this suggestion, larger "pica" type was used on the second preliminary edition and the final edition.

It was found that students had some difficulty if the question accompanying a particular drawing or description was on a different page. To compensate for this, questions on the second preliminary edition and the final edition were either on the same page as the drawing or description, or on the page directly opposite. This eliminated the flipping of pages back and forth to answer a question.

Second Pilot Test

Revised editions of Forms A and B were compiled on the basis of suggestions and item analyses data received on the first pilot test. The second pilot test was run during the last two weeks in February, 1972.

Sample

The second pilot test of revised Forms A and B of the Test of Science Inquiry Skills was administered to 185 students in the Van Dyke, Waterford, and East Lansing, Michigan, school districts, with the following breakdown of students in each district:

	<u>No. of Classrooms</u>	<u>No. of Students</u>
Van Dyke	2	50
Waterford	2	50
East Lansing	3	85

Students in the second pilot test were not involved in the first pilot test. Further, those East Lansing teachers involved in the second pilot test had also attended a summer SCIS workshop at Michigan State University in the summer of 1971.

As previously mentioned in the description of the first pilot test, Van Dyke students had been exposed to a textbook-oriented science program, while East Lansing students had been exposed to the SCIS program. Waterford students had been exposed to a traditional textbook-oriented course.

Purpose

The purpose of the second pilot testing program was to gather additional item analyses data. Many items on the second preliminary forms were revised on the basis of data gathered after the first pilot test.

Procedure

The procedure was basically similar to that followed in the first pilot testing program. Form A of the test was administered to Van Dyke and Waterford students, while Form B was administered to East Lansing students.

Results

Item analyses data obtained from the second pilot test were the major determining factor for selecting items to be included in the final version of the Test of Science Inquiry Skills. Forms A and B and summary data for each form are included in Appendix C.

Construction of Final Test

Length of Test

Based on feedback from both pilot testing programs, it was decided that 50 items would be a maximum number that students could reasonably be expected to complete in one hour.

Final Identification of Processes

With the limitation on the number of items on the test, it was decided that of the original list of processes identified in Figure 2, it would be more feasible to measure only four: (1) identifying and controlling variables, (2) interpreting data, (3) inferring, and (4) predicting, with approximately 12 test items in each category. These items would represent each subject matter area and grade level equally.

The final form of the instrument has the following item distribution grid in process and subject matter categories as shown in Figure 3.

Selection of the 50 items for the final test form was based on the following criteria: The item

- (1) had to be representative of one of the four process areas (identifying and controlling variables, interpreting data, inferring, and predicting);
- (2) had to be representative of one of the content areas of the SCIS program; and
- (3) had a minimum discrimination of 40 (except in cases where modifications were made to balance the test specification grid).

Approximately 140 items were field-tested, yet only 50 items representative of four process areas were included in the final edition. Because of the reduced number of process areas and overabundance of similar items in the

	N	Identifying & Controlling Variables (C) ^a	Interpreting Data (D)	Inferring (E)	Predicting (G)
(1) Populations	10		3	1	6
(2) Subsystems & Variables	8	5	3		
(3) Environments	10	3	2	3	2
(4) Relative Posi- tion & Motion	7		7		
(5) Communities	2			2	
(6) Energy Sources	13	1	3	6	3

Figure 3.--Final test specification grid.

^aRefer to Figure 2 for identification of process areas.

same process area, many items with acceptable indices of difficulty and discrimination were excluded from the final edition.

The Sample

The school districts of DeWitt, Grand Ledge, and Perry, Michigan, were selected because of their active participation with the SCIS program since its inception through the SCIS Trial Center at Michigan State University. The SCIS Trial Center has been in operation since 1967. The initial SCIS materials were used with the first grade and each year another grade was added. As SCIS was added to each grade, the textbook series (in use before SCIS was implemented) was dropped each year by that respective grade level. Thus, the fifth grade students used in this study, started the program in the first grade and have been exposed to each course in the appropriate sequence for the past five years. Each year that these students had the SCIS program, they were taught by teachers who were using SCIS materials for the first time at that particular grade level.

The following is a list of the SCIS schools utilized in the study (Experimental Group):

<u>Name of School</u>	<u>City</u>	<u>No. of Students</u>	<u>No. of Classrooms</u>
Fuerstenau	DeWitt	118	4
Delta Center	Grand Ledge	67	3
Perry	Perry	<u>152</u>	<u>5</u>
		337	12

Control schools were selected on the basis of results from the State of Michigan Assessment Examination, which is given to each fourth and seventh grade public school student in Michigan each year.

Previous studies that have utilized these SCIS schools have chosen control schools largely on the advice of local school administrators who were asked to identify schools they thought drew students from populations similar to their SCIS schools and had similar general educational programs with the exception of the SCIS science program.

It was thought that for this study, the utilization of State Assessment Examination scores would provide a better means of selecting schools that are similar to SCIS schools on measured factors.

The Michigan Educational Assessment Program gathers, analyzes, and reports three basic kinds of information descriptive of educational systems: (1) information regarding students' background characteristics; (2) information regarding school and school district educational resources (including data descriptive of finances, instructional staff, educational programs, and educational facilities); and (3) information regarding student/school performance (including data descriptive of attitudes and achievement in the basic skills).

This information is gathered from three basic sources: (1) an anonymous pupil background and attitude questionnaire administered to all fourth and seventh grade public

school students, (2) records held in the State of Michigan Department of Education, and (3) a basic skills achievement battery administered to all fourth and seventh grade public school students.

The 1970-71 fourth grade assessment results were used for selecting schools that were similar to the SCIS schools on measured factors. Control schools were chosen from the same geographic region¹ in the state (Region 2) and type of community where possible. Region 2 is an extremely large area. More specifically, all schools used in the study were relatively close to East Lansing and to each other. Schools used in the study were from the communities of Grand Ledge, DeWitt, Haslett, and Perry, Michigan. The first three are classified by the State Assessment program as "urban fringe," i.e., a community of any population size that has as its economic focal point a metropolitan core or a city. Perry is classified as a "town," which is characterized as a community of 2,500-10,000 that serves as the economic focal point of its environs.

Educational profiles for Michigan's geographic regions and community types include: (1) profiles constructed from district-level assessment results and (2) profiles constructed from school-level assessment results.

¹Michigan Department of Education, Local District Report: Explanatory Materials, Michigan Educational Assessment Report No. 6 (1970).

Control schools were thus selected from the same region in the state and type of community (as defined by the State Assessment) and had similar scores on: (1) student's estimate of socioeconomic status (SES), (2) performance on the basic skills measure of vocabulary, and (3) basic skills composite achievement.

The assessment battery which was given to students included several questions which were designed to gather information regarding students' socioeconomic background. The socioeconomic status measure is assumed to be indicative of students' perceptions of such things as the educational level of parents and their general economic level.²

To measure vocabulary, the assessment battery included 50 verbal analogy problems which measured students' knowledge of the meaning of words and their relationships.³

The composite achievement score was built by averaging the scores of the reading, English expression, and mathematics sections of the Assessment battery. The vocabulary score was not averaged into the composite achievement score.

Utilizing the above criteria, the following schools were selected as Control schools in the study:

²Ibid., p. 10.

³Michigan Department of Education, 1970-71 Individual Pupil Report: Explanatory Materials, Michigan Educational Assessment Program (April, 1971), p. 2.

<u>Name of School</u>	<u>City</u>	<u>No. of Students</u>	<u>No. of Classrooms</u>
Holbrook	Grand Ledge	61	2
Delta Mills	Grand Ledge	62	2
Murphy	Haslett	72	3
		195	7

Table 1 shows how these Control schools compared with their respective SCIS schools on scores of socioeconomic status, vocabulary, and composite achievement on the State Assessment examination.

Table 1.--Comparison of SCIS and Control schools.

School	<u>Socioeconomic Status</u>		<u>Vocabulary</u>		<u>Composite Achievement</u>	
	Mean	S.D. ^a	Mean	S.D.	Mean	S.D.
Fuerstenau (SCIS)	52.5	7.7	51.0	9.7	50.6	8.6
Holbrook (Control)	53.2	8.1	51.9	10.6	50.9	9.2
Delta Center (SCIS)	57.2	6.6	52.5	9.0	51.1	8.8
Delta Mills (Control)	54.1	9.8	53.6	7.4	52.4	7.8
Perry (SCIS)	50.1	9.1	48.9	8.9	46.9	9.0
Murphy (Control)	51.1	10.5	48.3	9.0	46.9	8.7

^aStandard deviation.

Use of Other Instruments in the Study

Science and Scientists Attitude Inventory (SASAI)

One of the goals of the SCIS program is to improve the student's attitude toward science and to create a more realistic attitude toward science and scientists. This inventory, developed by Motz, was utilized to obtain a more balanced assessment of the effect of the SCIS program.⁴ It contains ideas and statements about science and scientists that were obtained by questioning elementary, secondary, and college students, plus scientists and science educators. The student reacts to each statement by either agreeing, disagreeing, or remaining undecided.

State of Michigan Assessment Composite Achievement Score

As previously mentioned, a composite achievement score was obtained by averaging the individual's standard scores on the reading, the mechanics of written English, and the mathematics tests.⁵ The test scores were averaged in such a way that each score contributed equally to the average, despite the fact that the number of items was different on the three tests.

⁴ LaMoine L. Motz, "The Development of an Instrument to Evaluate Sixth and Ninth Grade Students' Attitudes Toward Science and Scientists" (unpublished Ph.D. dissertation, University of Michigan, 1970).

⁵ Michigan Department of Education, 1970-71 Individual Pupil Report: Explanatory Materials, p. 4.

The reading test contained 50 questions which assessed paragraph comprehension, ability to understand words from the context in which they are encountered, and ability to identify the correct synonym for a word. Students were allowed 35 minutes to work on this test.⁶

The mechanics of written English test consisted of four parts, each separately timed. In part A, spelling, students were to identify misspelled words. The fourth grade test presented 15 items to be completed in five minutes. Part B, effectiveness of written expression, contained 14 items and nine minutes were allowed for its completion. Students were required to select the best way of expressing a thought. Eight minutes were allowed for part C, written usage. Fourth grade students were to recognize grammatical errors in 14 items. To recognize errors of punctuation and capitalization was the object of part D, punctuation and capitalization. Eight minutes were allowed for part D, which contained 12 items.⁷

The mathematics test involved mathematical reasoning and problem solving. Thirty minutes were allowed to answer 40 questions.

⁶Ibid., p. 3.

⁷Ibid.

Sequential Test of Educational
Progress (STEP) Series II

This evaluative instrument is a nationally normed, standardized achievement test published by the Educational Testing Service. The Science Test (Form 4A) was the test utilized in this study. While its main emphasis is on content areas, science processes are also included on the test. The utilization of this test allowed for comparisons of the students' results on the Test of Science Inquiry Skills with a nationally normed test. It is a well-constructed test with a reliability of 0.86. The validity of the test was not mentioned in the SCAT-STEP Series II Book of Norms,⁸ nor in the test manual. The test emphasizes application and understanding rather than factual recall. Much use is made of the interpretive exercise.⁹

Further, the STEP test was utilized to help in determining if those students exposed to the SCIS program, which dually emphasizes content areas and processes, were able to do as well on an achievement test such as the STEP test, as non-SCIS students, even though it was written mainly for evaluating content-oriented programs.

⁸Cooperative Tests and Services, SCAT-STEP, Series II Book of Norms (Princeton, New Jersey: Educational Testing Service, 1970).

⁹William A. Mehrens and Irvin J. Lehmann, Standardized Tests in Education (New York: Holt, Rinehart and Winston, Inc., 1969), pp. 165-67.

Design of Study

It is believed that the static-group comparison¹⁰ accurately depicts the design of the study:

$$\frac{X}{0_1} \quad - \quad - \quad \frac{0_2}{0_2}$$

This is a design in which a group of fifth grade students which has experienced X (SCIS for five years) is compared with one which has not, for the purposes of establishing the effect of X.

There is no formal means of certifying that the groups would have been equivalent had it not been for X. 0_1 and 0_2 refer to the battery of tests which was administered to both groups.

The general design of the study is depicted in Figure 4.

The experimental unit in this study is the classroom. However, some consideration must be given to the fact that the students have not always been in the same classroom group throughout the five years of the SCIS program, and have not had the same teachers.

The repeated measures (split plot) analysis was used in analyzing the data because it provides for tests of the treatment main effect and treatment by measures interaction. If treatment by measures interaction is indicated, then post

¹⁰Donald T. Campbell and Julian C. Stanley, Experimental and Quasi-Experimental Designs for Research (Chicago: Rand McNally and Company, 1963), p. 12.

hoc procedures (utilizing a series of one-way anova) can be used to determine if the treatment effects are measure specific. As previously mentioned in Chapter II, there is the possibility of some treatment by measures interaction.

	SASAI	TSIS	STEP	STATE ASSESSMENT
SCIS Classrooms 1 - 12				
Control Classrooms 13 - 19				

Figure 4.--General design of study.

To determine the strength of relationship between the instruments utilized in this study, a series of correlation coefficients between the tests used is also calculated.

Final Testing Program

Contact With Schools

The principals of all schools selected for the study were contacted to obtain permission to utilize various classes within the school. It was made clear that the individual

names of the students were not to be made public; however, the results of each classroom's performance on the tests were to be made available to those teachers who so desired that information.

Further, each principal in the SCIS schools was asked to identify the number of years each student had been exposed to the SCIS program. For example, if the student had entered the school in the fourth grade, he would have been exposed to the program for only two years. The length of time the students were in the program was, therefore, established.

Table 2 shows the distribution, by number of years in the SCIS program, of all SCIS students who participated in the final testing program. Only those students who had been in the SCIS program for five years were used in the study.

Table 2.--Years of students' SCIS experience in Experimental classrooms.

School District	Classroom	No. of Students	Years in SCIS Program		
			5	3-4 ^a	1-2 ^a
DeWitt	1	29	22	1	6
	2	29	21	6	2
	3	31	14	6	11
	4	29	13	7	9
Grand Ledge	5	21	9	5	7
	6	23	14	9	0
	7	23	11	7	5
Perry	8	31	9	17	5
	9	29	14	8	7
	10	31	21	5	5
	11	30	22	3	5
	12	31	23	4	4
TOTALS		337	193	78	66

^aNot utilized in study.

Administration of Tests

The instruments SASAI, TSIS, and STEP were administered to the previously mentioned schools during the two-week interval between May 8-22, 1972. Instructions for administering the tests were provided for each participating teacher (and are included in the appendix).

On May 22, 1972, all test booklets and completed answer sheets were picked up from the schools. These answer sheets were then machine scored and subjected to item analyses by the Evaluation Services at Michigan State University.

Thank you letters as well as test results were later sent to each principal.

Item Analyses Used in the Study

Standard item analyses measures were used in evaluating items throughout the study.

Item Difficulty

Item difficulty is reported as the proportion of the total group that got the item wrong; thus, a high index indicates a difficult item, and a low index indicates an easy item.

Item Discrimination

Item discrimination is calculated as the difference between the proportion of the upper 27 per cent of the group who got the item correct and the proportion of the lower 27 per cent who got the item correct.

Figure 5 is included to help clarify the following discussion of item difficulty and item discrimination indices.

The regular form shows the item number; key (number of the correct option); the percentage of the upper, middle, and lower groups who selected each option; the index of difficulty; and the index of discrimination. For example, in item 1 in Figure 5, option 4 was the correct answer and it was selected by 64 per cent of the upper group, 62 per cent of the middle group, and 0 per cent of the lower group. The index of difficulty, based on the total group, was 54 and the index of discrimination was 64.

The final set of items was evaluated using reliability and standard error measurement indices.

Reliability

The reliability (internal consistency reliability coefficient) is estimated by the Kuder-Richardson Formula #20. The estimate is an indication of the homogeneity of the items in the test, or the degree to which the item responses correlate with the total test score.

Standard Error of Measurement

Standard Error of Measurement is based on the standard deviation of the test scores and on the estimated reliability of the test. The final version of the Test of Science Inquiry Skills given to 310 students from SCIS schools produced the following results:

Item Analysis		Test 4482		125 items		112 students	
		Percentages		Upper 27%		Lower 27%	
<u>Item</u>	<u>Key</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>Disc</u>
1	4	7-23-57	0-4-7	29-8-36	64-62-0	0-4-0	64
2	2	2-12-7	64-42-29	14-4-21	14-42-36	0-0-7	35

Figure 5.--Sample of item analyses data sheet.

Mean item difficulty: 46

Mean item discrimination: 49

Kuder-Richardson Formula #20 Reliability: 0.9033

Standard Error of Measurement: 3.0810

More detailed results are included in the appendix.

The final form of the test was designed to measure four processes: (1) identifying and controlling variables, (2) interpreting data, (3) predicting, and (4) inferring. All the items designed to measure each of these processes were defined as a subtest and subjected to the same item analyses as the entire test. Table 3 includes a listing of the four subtests and the test items in each subtest.

Table 4, the final test specification grid, summarizes the process and content areas included on the final form of the Test of Science Inquiry Skills and the test items representative of each category.

Summary

The following steps were taken in developing the Test of Science Inquiry Skills:

- (1) Identification of the process and concept areas stressed in the SCIS program in grades 3, 4, and 5.
- (2) Statements of behavioral objectives designed to indicate what the student must do if and when the particular objective has been met.
- (3) Construction of a grid with concept objectives on the vertical axis and processes on the horizontal axis.
- (4) Writing of test items within the confines of the grid to measure the identified processes within the prescribed subject-matter areas.

Table 3.--Grouping of test items into subtests on the basis of process.

Identifying & Controlling Variables		Interpreting Data	Inferring	Predicting
Item #	7	10	1	2
	13	11	6	3
	19	12	8	4
	25	15	14	5
	29	16	17	9
	30	21	18	24
	39	22	20	28
	40	23	26	31
	41	27	38	32
		33	46	48
		34	47	49
		35	50	
		36		
		37		
		42		
		43		
		44		
		45		
TOTALS	9	18	12	11

Table 4.--Final test specification grid.

	(C) Identifying and Controlling Variables	(D) Interpreting Data	(E) Inferring	(G) Predicting
1.00 Population		21,22,23		
1.1 Predator-prey				2,3,4,5
1.2 Food chain			1	24
1.3 Populations interdependence				49
2.00 Subsystems & Variables				
2.1 Evidence of interaction	7,13,39, 40,41			
2.2 Identifying systems		10,11,12		
2.21 Isolating subsystems				
3.00 Environments		45	46,47	
3.1 Environmental factor	25			
3.2 Range of environmental factor	29,30	27	26	
3.3 Optimum range				28,48
4.00 Relative Position & Motion		33,34,35, 36,42,43, 44		
5.00 Communities				
5.1 Plant eater-animal eater			50	
5.4 Cyclic path of materials			6	
6.00 Energy Sources	8	37	18,38	
6.1 Energy transfer from energy source to energy receiver		15,16	14,17,20	31,32
6.2 Variables which affect energy transfer	19			9

- (5) Pilot testing of items on a small sample of students to detect extreme reading and mechanical problems with the items.
- (6) Correction and arrangement of items into two tests, Forms A and B, each with approximately 70 separate items.
- (7) Administration of each form to 900 students in three different school districts.
- (8) Subjection of student responses from the pilot testing procedure in (7) above to item analyses.
- (9) Elimination and/or revision of test items; completion of revised versions of Forms A and B.
- (10) Further pilot testing of revised versions of Forms A and B with 100 additional students for each form.
- (11) Subjection of results to item analyses.
- (12) Compilation of final 50-item single test, based on item analyses of the second pilot test and the degree to which these items completed the selected areas on the grid mentioned in (3) above.

The following steps were taken in the study:

- (1) Selection of experimental and control groups based on similarities of socioeconomic and composite achievement scores on the State of Michigan Assessment Examination.
- (2) Selection of other instruments (SASAI, STEP, State Assessment composite achievement score) for use in the study.
- (3) Administration of the four instruments to 310 SCIS students who had been in the program for five consecutive years, and 195 Control students.
- (4) Analyses of data from all four instruments for both SCIS and Control groups based on the statistical design of the study (static-group comparison).

CHAPTER IV

ANALYSIS AND INTERPRETATION OF THE DATA

One purpose of this study was to develop a paper and pencil test for fifth grade students based on the stated or implied goals for grades 3, 4, and 5 of the Science Curriculum Improvement Study (SCIS). The four processes of identifying and controlling variables, interpreting data, inferring, and predicting were identified as part of the SCIS curriculum at these three grade levels.

A test consisting of 50 multiple choice items was developed and administered to a sample of 310 fifth grade SCIS students, and 191 fifth grade Control students. Statistical analyses of the resulting data are presented in this chapter.

Selection of the 50 items for the final test form was based on the following criteria. Each item selected:

- (1) was representative of one of the four process areas (identifying and controlling variables, interpreting data, inferring, and predicting);
- (2) was representative of one of the content areas of the SCIS program; and

- (3) had a minimum discrimination of 40 (except in some special cases where modifications were made to balance the test specification grid).

The final test, designed for fifth grade students, included items equally representative of all three grade level content and process areas.

Test Item Analyses

The item analyses for instruments used in this study were computed at the Evaluation Services of Michigan State University. The complete data analyses for the test items are presented in Table 5.

The item analyses data in Table 5 contain four kinds of information for each of the items:

- (1) the keyed correct response
- (2) the content and process area that each item measures
- (3) the item difficulty: this value as recorded is the percentage of students in the total group who incorrectly answered the particular item. For example, if 45 students among the 60 tested responded correctly, then 15 incorrect responses correspond to 25 per cent of the students. The item difficulty in this case would be recorded as 25. An item with a high item difficulty index was answered incorrectly by a high percentage of students.
- (4) item discrimination reported as a per cent value: this is calculated by subtracting the percentage of

Table 5.--Summary data on items for Test of Science Inquiry Skills (SCIS population).

Item #	Correct Response	Process and Content Area ^a	Difficulty	Discrimination
1	2	E, 1.2	35	32
2	4	G, 1.1	16	34
3	2	G, 1.1	45	34
4	2	G, 1.1	45	51
5	3	G, 1.1	20	29
6	1	E, 5.4	43	56
7	2	C, 2.21	56	30
8	4	E, 6.0	26	47
9	3	G, 6.2	43	51
10	3	D, 2.2	47	41
11	1	D, 2.2	60	53
12	3	D, 2.2	55	43
13	2	C, 2.21	62	18
14	1	E, 6.1	53	60
15	4	D, 6.1	23	55
16	4	D, 6.1	33	61
17	2	E, 6.1	44	65
18	1	E, 6.0	38	65
19	3	C, 6.2	49	74
20	2	E, 6.1	44	60
21	4	D, 1.0	32	59
22	2	D, 1.0	28	66
23	2	D, 1.0	49	54
24	1	G, 1.2	55	30
25	1	C, 3.1	42	57
26	2	E, 3.2	48	57
27	3	D, 3.2	50	28
28	4	G, 3.3	60	29
29	2	C, 3.2	49	62

Table 5.--Continued.

Item #	Correct Response	Process and Content Area ^a	Difficulty	Discrimination
30	4	C, 3.2	55	37
31	2	G, 6.1	54	41
32	1	G, 6.1	63	35
33	3	D, 4.0	35	60
34	4	D, 4.0	33	52
35	1	D, 4.0	37	70
36	3	D, 4.0	30	64
37	2	D, 6.0	33	64
38	1	E, 6.0	35	69
39	2	C, 2.21	45	60
40	1	C, 2.21	39	81
41	3	C, 2.21	38	64
42	1	D, 4.0	48	28
43	3	D, 4.0	26	56
44	1	D, 4.0	80	9
45	2	D, 3.0	48	52
46	3	E, 3.0	41	75
47	1	E, 3.0	56	68
48	1	G, 3.3	45	61
49	1	G, 1.3	60	42
50	3	E, 5.1	56	70

^aSee item specification grid, Table 4.

correct responses among students who were in the lower 27 per cent on the entire test from the percentage of students who were in the upper 27 per cent of the group. For example: if 15 out of 20 students in the upper group, and 10 out of 20 students in the lower

group got the item correct, the item discrimination would be 0.75-0.50, or 0.25.

Standardization

Standard scores with a mean of 50 and a standard deviation of 10 were calculated for SCIS and Control populations for each of the four tests. These are included in the appendix.

Reliability

The reliability for the Test of Science Inquiry Skills (TSIS) was determined by the Kuder-Richardson Formula #20. The total instrument (50 items) reliability was 0.90 for the SCIS population and 0.87 for the Control population.

Correlations between alternate forms of the test and test-retest correlations have not been attempted.

Teachers were requested to allow enough time for their students to complete the test. Feedback from teachers indicated that one hour provided students ample time to complete all items. Also, in reviewing the answer sheets, it was evident that the number of items omitted did not increase appreciably at the end of the test.

All the items in the test which were identified as measuring a given process were defined as a subtest and a Kuder-Richardson Formula #20 reliability coefficient was calculated for each of these subtests. The hypothesis that the reliability coefficient for the entire test of 50 items testing various processes was the same as that for a subset

of items professing to test one process, was tested for each subset. If the reliability coefficient for the subset of items was significantly higher than for the whole test, items in the subtest would have been shown to be more highly correlated than those of the whole test.

A correction formula for estimating increased reliability with increased test length was applied to estimate what the reliability of the subtests would be if they were the same length as the original test. The formula¹ is as follows:

$$r_{xx} = \frac{Kr}{1 + (K - 1)r}$$

where r_{xx} = the predicted reliability of a test K times as long as the original test

r = the reliability of the original test

K = the ratio of the number of items in the new test to the number of items in the original one.

Table 6 summarizes these reliabilities.

It should be noted that the extremely high Kuder-Richardson Formula #20 coefficient of 0.9033 for the entire test for the SCIS population indicates that the items are very homogeneous and thus it would be extremely difficult to obtain higher coefficients for the subtests. The Kuder-

¹William A. Mehrens and Irvin J. Lehmann, Standardized Tests in Education (New York: Holt, Rinehart and Winston, Inc., 1969), p. 39.

Richardson Formula #20 (corrected for length) for "inferring," "identifying and controlling variables," and "interpreting data" are numerically higher than the KR #20 for the entire test.

Table 6.--Summaries of reliability for SCIS and Control populations; and for subtests with SCIS population only.

	# of Items	Index of Diff.	Index of Disc.	Stand. Error	Calc. KR 20	KR 20 Corr. for Length
Control	50	39	43	3.1017	0.8726	
SCIS	50	46	49	3.0810	0.9033	
<u>Subtests^a</u>						
Identifying & Controlling Variables	9	48	61	1.3511	0.6394	0.9097
Interpreting Data	18	41	55	1.8176	0.7935	0.9154
Inferring	12	42	66	1.4644	0.7865	0.9385
Predicting	11	46	51	1.4901	0.5728	0.8589

^aSCIS population only.

The individual subtests consist of items identified as measuring a single process. No attempt was made to balance content area representation in each subtest. The science content area was utilized only as a framework in which a process skill was tested. As previously mentioned, process items do not require the student to bring factual information with him to the test.

Data for each subtest treated as a distinct test are summarized in Table 6. Tables 7, 8, 9, and 10 provide a more detailed review of the item analyses for each subtest treated as a complete test.

Validity

A critical question of an instrument such as the Test of Science Inquiry Skills is its validity. There is considerable evidence for the content validity of this test, due to the care employed in the construction and validation of the two-dimensional test specification grid and the development of items from this grid. Science educators familiar with the process and content goals of the SCIS program were consulted during the construction of the grid and the development and tryout of the items. (A complete discussion of grid development and item development was presented in Chapter III.)

The criterion related validity of this test is very difficult to investigate because of the limited number of appropriate external criterion measures available that evaluate science process skills of children at the fifth grade level. The AAAS program Science--A Process Approach has developed a series of process competency measures for grades 4, 5, and 6 that can be administered to individuals or groups of students, depending on which form is used. As previously mentioned in Chapter II, it was thought that the group competency measure for fifth grade students would be

Table 7.--Summary data for subtest Identifying and Controlling Variables.

<u>Item #</u>	<u>Correct Response</u>	<u>Content Area^a</u>	<u>Difficulty</u>	<u>Discrimination</u>
7	2	2.1	56	46
13	2	2.21	62	36
19	3	6.2	49	75
25	1	3.1	41	65
29	2	3.2	49	69
30	4	3.2	55	51
39	2	2.21	45	65
40	1	2.21	39	77
41	3	2.21	38	70

DISTRIBUTION OF ITEM DIFFICULTY INDICES

	<u>No. of Items</u>	<u>Percentage</u>
91-100		
81- 90		
71- 80		
61- 70	1	11
51- 60	2	22
41- 50	4	44
31- 40	2	22
21- 30		
11- 20		
0- 10		

DISTRIBUTION OF DISCRIMINATION INDICES

	<u>No. of Items</u>	<u>Percentage</u>
91-100		
81- 90		
71- 80	2	22
61- 70	4	44
51- 60	1	11
41- 50	1	11
31- 40	1	11
21- 30		
11- 20		
0- 10		

^aSee item specification grid, Table 4, for key to content areas.

Table 8.--Summary data for subtest Interpreting Data.

<u>Item #</u>	<u>Correct Response</u>	<u>Content Area^a</u>	<u>Difficulty</u>	<u>Discrimination</u>
10	3	2.2	47	52
11	1	2.2	59	68
12	3	2.2	55	52
15	4	6.1	23	59
16	4	6.1	33	60
21	4	1.0	32	60
22	2	1.0	28	69
23	2	1.0	48	59
27	3	3.2	50	36
33	3	4.0	34	65
34	4	4.0	33	52
35	1	4.0	37	77
36	3	4.0	30	71
37	2	6.0	33	66
42	1	4.0	48	29
43	3	4.0	26	58
44	1	4.0	80	12
45	2	3.0	48	49

DISTRIBUTION OF ITEM DIFFICULTY INDICES

	<u># of Items</u>	<u>%</u>		<u># of Items</u>	<u>%</u>
91-100			41- 50	5	28
81- 90			31- 40	6	33
71- 80	1	6	21- 30	4	22
61- 70			11- 20		
51- 60	2	11	0- 10		

DISTRIBUTION OF DISCRIMINATION INDICES

	<u># of Items</u>	<u>%</u>		<u># of Items</u>	<u>%</u>
91-100			41- 50	1	6
81- 90			31- 40	1	6
71- 80	2	11	21- 30	1	6
61- 70	4	22	11- 20	1	1
51- 60	8	44	0- 10		

^aSee item specification grid, Table 4, for key to content areas.

Table 9.--Summary data for subtest Predicting.

<u>Item #</u>	<u>Correct Response</u>	<u>Content Area^a</u>	<u>Difficulty</u>	<u>Discrimination</u>
2	4	1.1	16	38
3	2	1.4	46	54
4	2	1.1	45	67
5	3	1.1	20	32
9	3	6.2	43	57
24	1	1.2	55	51
28	4	3.3	60	45
31	2	6.1	54	45
32	1	6.1	63	51
48	1	3.3	45	71
49	1	1.3	60	55

DISTRIBUTION OF ITEM DIFFICULTY INDICES

	<u>No. of Items</u>	<u>Percentage</u>
91-100		
81- 90		
71- 80		
61- 70	1	9
51- 60	4	36
41- 50	4	36
31- 40		
21- 30		
11- 20	2	18
0- 10		

DISTRIBUTION OF DISCRIMINATION INDICES

	<u>No. of Items</u>	<u>Percentage</u>
91-100		
81- 90		
71- 80	1	9
61- 70	1	9
51- 60	5	45
41- 50	2	18
31- 40	2	18
21- 30		
11- 20		
0- 10		

^aSee item specification grid, Table 4, for content areas.

Table 10.--Summary data for subtest Inferring.

<u>Item #</u>	<u>Correct Response</u>	<u>Content Area^a</u>	<u>Difficulty</u>	<u>Discrimination</u>
1	2	1.2	24	42
6	1	5.4	43	59
8	4	6.0	26	52
14	1	6.1	53	69
17	2	6.1	44	73
18	1	6.0	39	75
20	2	6.1	44	68
26	2	3.2	48	66
38	1	6.0	35	74
46	3	3.0	41	78
47	1	3.0	56	71
50	3	5.1	56	72

DISTRIBUTION OF ITEM DIFFICULTY INDICES

	<u>No. of Items</u>	<u>Percentage</u>
91-100		
81- 90		
71- 80		
61- 70		
51- 60	3	25
41- 50	5	42
31- 40	2	17
21- 30	2	17
11- 20		
0- 10		

DISTRIBUTION OF DISCRIMINATION INDICES

	<u>No. of Items</u>	<u>Percentage</u>
91-100		
81- 90		
71- 80	6	50
61- 70	3	25
51- 60	2	17
41- 50	1	8
31- 40		
21- 30		
11- 20		
0- 10		

^aSee item specification grid, Table 4, for content areas.

difficult for the teachers to administer. Also, the answer sheets are not machine scorable.

Further, a study conducted by Walbesser and Carter² to determine if the AAAS competency measures for levels E and F (grades 4 and 5) could be converted to group competency measures indicated that the group format produced consistently lower scores than the individual measures on the four integrated processes of controlling variables, defining operationally, formulating hypotheses, and interpreting data. These are some of the reasons why the AAAS group competency measures for fifth grade students were not utilized as a criterion measure in this study.

In an attempt to provide at least some evidence of criterion related validity, students who took the Test of Science Inquiry Skills also took the Science Test (Form 4A) from the Sequential Test of Educational Progress (STEP) Series II testing battery. This test is designed to measure knowledge of important topics in science, comprehension and application of this knowledge, and mastery of science skills.³ Of the nationally normed achievement tests available, this test has more process-type items than any other.

²Henry H. Walbesser and Heather L. Carter, "The Effect on Test Results of Changes in Task and Response Format Required by Altering the Test Administration From an Individual to a Group Form," Journal of Research in Science Teaching, VII, 1 (1970), 1-8.

³Cooperative Tests and Services, Bulletin of Tests and Services (Princeton, New Jersey: Educational Testing Service, 1971), p. 12.

In reviewing the STEP test, the following items were identified as process items:

<u>Process</u>	<u>Item Number(s)</u>	<u>Total</u>
Interpreting data*	3, 12, 38, 43	4
Controlling variables*	4	1
Predicting*	29, 31	2
Inferring*	10, 17	2
Experimental design	7	1
Application	25	<u>1</u>
		11

*Also found in TSIS.

Four of the six processes tested are the same as those included in the Test of Science Inquiry Skills. Also, nine of the 11 process items measure the same processes measured in the TSIS.

While nine out of 50 items is not a large proportion, at least the tests have some similarities--perhaps more than would be found with any other standardized achievement test.

As part of the major design of the study, all children took four tests: the TSIS, the SASAI, the STEP (Series II, Form 4A Science Test), and the composite achievement of the State of Michigan Assessment Examination. The State Assessment Reading score was also available for each student. Classroom and school means on each of the four tests are summarized in Table 11. Correlation coefficients were calculated between all five test scores available. All

Table 11.--Summary of classroom and school means comparison by SCIS and Control schools.

SCHOOL AND CLASSROOM #	SASAI		TSIS		STEP		STATE ASSESS.		READING	
	C.R. $\bar{X}$	SCH. $\bar{X}$	C.R. $\bar{X}$	SCH. $\bar{X}$	C.R. $\bar{X}$	SCH. $\bar{X}$	C.R. $\bar{X}$	SCH. $\bar{X}$	C.R. $\bar{X}$	SCH. $\bar{X}$
Fuerstenau (SCIS)		50.51		49.22		49.84		50.22		51.89
1	49.72		52.04		51.12		49.00		50.28	
2	49.32		43.55		50.07		50.72		52.66	
3	49.29		48.47		48.09		50.16		50.75	
4	53.71		52.80		50.07		51.00		53.90	
Holbrook (Control)		47.41		51.79		47.97		51.26		50.94
13	46.34		48.82		45.88		51.96		51.50	
14	48.47		54.77		50.06		50.57		50.39	
Delta Center (SCIS)		51.15		52.17		49.22		50.19		48.66
5	53.16		50.91		47.33		48.62		49.50	
6	46.95		53.54		48.65		49.25		47.58	
7	53.31		52.08		51.67		52.70		48.90	
Delta Mills (Control)		48.31		53.69		52.94		52.01		53.23
15	50.91		54.37		53.63		54.03		54.17	
16	45.72		53.01		52.26		50.00		52.30	
Perry (SCIS)		48.47		47.27		49.20		48.11		47.58
8	43.63		42.27		46.01		45.11		46.55	
9	49.82		50.55		48.09		47.71		46.00	
10	47.91		47.46		51.47		48.00		47.35	
11	52.61		46.42		51.56		48.25		46.10	
12	48.37		51.67		48.87		51.50		51.90	
Murphy (Control)		52.63		47.76		48.67		45.67		44.73
17	55.17		49.42		48.18		43.70		44.10	
18	49.17		47.18		49.65		47.47		45.62	
19	53.56		52.68		48.20		45.84		44.47	

correlations were significantly different from zero at the 0.001 level of significance except between the SASAI and Reading, which were significant at the 0.01 level. Figure 6 shows these correlations.

SASAI					
TSIS	.4358 ^a .3837 ^b				
STEP	.3968 ^a .4072 ^b	.6723 ^a .6393 ^b			
STATE ASSESS.	.3202 ^a .3921 ^b	.5563 ^a .6266 ^b	.5624 ^a .6709 ^b		
READING	.2983 ^a .3501 ^b	.4989 ^a .6172 ^b	.5344 ^a .6605 ^b	.8938 ^a .9520 ^b	
	SASAI	TSIS	STEP	STATE ASSESS.	READING

^aSCIS
^bControl

Figure 6.--Correlation coefficients between test scores.

As predicted, student scores on the STEP and TSIS show a high correlation. The relatively high correlation between the TSIS and reading was also predictable based on feedback from teachers and students throughout the development of the instrument.

The relatively high correlation between the STEP and reading and between the TSIS and reading indicates that perhaps both of these tests favor good readers.

Since one-third of the State Assessment composite achievement score is based on reading, perhaps the high correlation of the STEP and TSIS to the State Assessment score could be related to reading ability. The SASAI has a low correlation with reading and also a low correlation with the TSIS and STEP tests.

Hypotheses

As previously mentioned, the purpose of this study was to develop a process instrument and to use this and other instruments to estimate to what degree the SCIS program has met some of its stated or implied goals. The most important hypotheses tested were as follows:

H_{o_1} : There is no significant difference between classrooms receiving science instruction with SCIS and classrooms not receiving SCIS instruction as measured by the Science and Scientists Attitude Inventory (SASAI), the Test of Science Inquiry Skills (TSIS), the Sequential Test of Educational Progress (STEP), and the State of Michigan Assessment.

H_{o_2} : On the SASAI, TSIS, STEP, and State of Michigan Assessment, there will be no significant interaction between treatment and measures.

Ho₃: There is no significant difference between classrooms receiving SCIS science instruction and Control classrooms when comparing their achievement as measured by the State of Michigan Assessment composite achievement score.

Ho₄: There is no significant difference between classrooms receiving SCIS instruction and Control classrooms when comparing their ability to solve science problems as measured by the TSIS.

Ho₅: There is no significant difference between classrooms receiving SCIS instruction and Control classrooms when comparing their ability to recall factual science information as measured by the STEP.

Ho₆: There is no significant difference between classrooms receiving SCIS instruction and Control classrooms when comparing their attitude toward science and scientists as measured by the SASAI.

Design of the Study

It is believed that the static-group comparison⁴ accurately depicts the design of the study:

$$\frac{X}{\quad} \quad - - \quad \frac{O_1}{O_2}$$

This is a design in which a group of fifth grade students which has experienced X (SCIS for five years) is compared

⁴Donald T. Campbell and Julian C. Stanley, Experimental and Quasi-Experimental Designs for Research (Chicago: Rand McNally & Company, 1963), p. 12.

with one which has not, for the purposes of establishing the effect of X.

Since it was not possible to randomly assign students to SCIS and Control groups, the study does not include an adequate Control group. In an effort to find a Control group, State Assessment scores were utilized to identify those schools similar to the SCIS schools.

O_1 and O_2 refer to the battery of tests which was administered to each group.

To better visualize the general design of the study, Figure 7 has been included. Twelve classrooms which had received SCIS instruction and seven classrooms which had no SCIS instruction took all four tests.

	SASAI	TSIS	STEP	STATE ASSESSMENT
SCIS Classrooms 1 - 12				
Control Classrooms 13 - 19				

Figure 7.--General design of study.

Repeated Measures

In this study the general questions of interest include: Is there an overall treatment effect as measured by the Science and Scientists Attitude Inventory, the Test of Science Inquiry Skills, the Sequential Test of Educational Progress, and the State of Michigan Assessment Examination; and is there a treatment by measures interaction (does one treatment group behave differently on one or more of the tests)? Since the data consist of four different test scores, each test score was transformed into a standard score with a mean of 50 and a standard deviation of 10. This transformation into a common metric was done to facilitate interpretation of the treatment by measures interaction.

Since the children involved in this study have received their science instruction as a classroom unit, the statistical unit used is the classroom. Classroom averages for each of the four instruments were calculated using transformed scores of students who had been in the SCIS program for all five years. Students who had been in the SCIS program for less than five years were not included in the study. Table 12 shows the class mean on each of the four instruments utilized in the study.

The repeated measures analysis of variance was calculated on the transformed scores. The results are summarized in Table 13.

The critical value for testing for the treatment main effect at $\alpha = 0.05$ with 1, 17 df is 4.45. Since the

Table 12.--Summary of the means on the four tests by classroom.

Classroom	School	SASAI	TSIS	STEP	STATE ASSESSMENT ^a
SCIS					
1	DeWitt	49.72	52.04	51.12	49.00
2		49.32	43.55	50.07	50.72
3		49.29	48.47	48.09	50.16
4		53.71	52.80	50.07	51.00
5	Delta Center	53.16	50.91	47.33	48.62
6		46.95	53.54	48.65	49.25
7		53.31	52.08	51.67	52.70
8	Perry	43.63	42.27	46.01	45.11
9		49.82	50.55	48.09	47.71
10		47.91	47.46	51.47	48.00
11		52.61	46.42	51.56	48.25
12		48.37	51.67	48.87	51.50
CONTROL					
13	Holbrook	46.34	48.82	45.88	51.96
14		48.47	54.77	50.06	50.57
15	Delta Mills	50.91	54.37	53.63	54.03
16		45.72	53.01	52.26	50.00
17	Murphy	55.17	49.42	48.18	43.70
18		49.17	47.18	49.65	47.47
19		53.56	52.68	48.20	45.84

^aComposite achievement score.

calculated value of the test statistic is 0.313, we fail to reject the null hypothesis of no treatment main effect.

Table 13.--Results of repeated measures by the use of analysis of variance.

Source of Variation	df	Sum of Squares	Mean Squares	F	P
Groups	1	4.8968	4.8968	0.313	N.S.
Subj-G.	17	265.6716	15.6277		
Rep. Measures	3	7.6273	2.5424	0.405	N.S.
RM*G	3	15.9374	5.3124	0.846	N.S.
RM*S-G	51	320.2158	6.2787		
TOTAL	75	614.3492			

The test of treatment by measures interaction requires that the measures be equally correlated among themselves. To avoid this assumption a conservative test⁵ is used. This test defines the region of rejection by reducing the degrees of freedom.

The critical value for testing the treatment by measures interaction using the conservative test with $\alpha = 0.05$ with 1, 17 degrees of freedom is 4.45. Since the calculated value of the test statistic is 0.846, we fail to reject the conservative test. Even with the liberal test with 3, 51 df, we fail to reject the hypothesis of interaction between

⁵Roger E. Kirk, Experimental Design: Procedures for the Behavioral Sciences (Belmont, California: Brooks/Cole Publishing Co., 1968).

treatment and measures. Since there is no interaction, no post hoc procedures for simple main effects were necessary.

To better understand the data, a graph of the treatment by measures data has been constructed, with measures on the ordinate and standard scores on the abscissa, and is found in Figure 8.

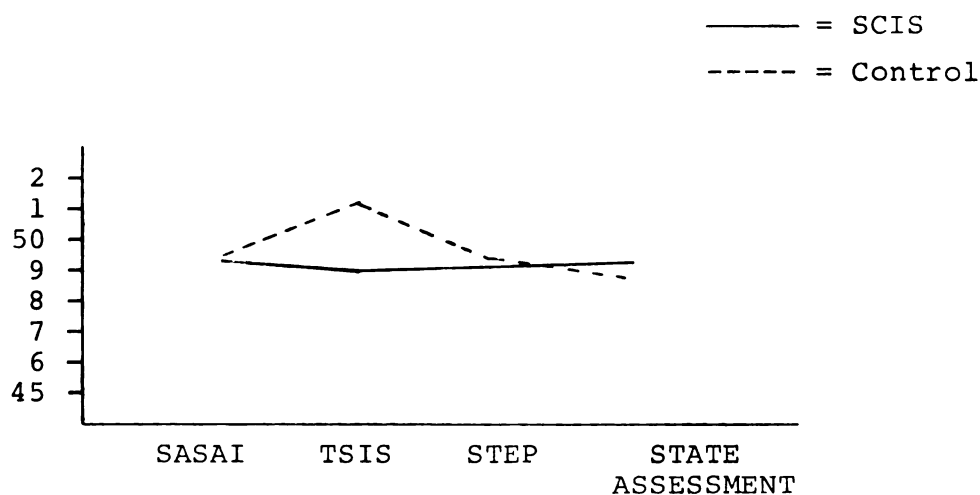


Figure 8.--Summary of treatment by measures data.

The following table of classroom mean scores summarizes the above graph:

	SASAI	TSIS	STEP	STATE ASSESSMENT
SCIS	49.80	49.31	49.58	49.33
Control	49.90	51.46	49.69	49.08

Results

We fail to reject Hypotheses 1 and 2, that there is no treatment main effect and interaction. Since there is no

significant interaction, we also fail to reject Hypotheses 3, 4, 5, and 6.

With the student used as the experimental unit, significant interaction was approached at the 0.25 level of significance.

Summary

The following summarizes the methods used in analyzing the data:

- (1) Item difficulty and discrimination were calculated on each test item.
- (2) All scores were converted into standard scores with a mean of 50 and a standard deviation of 10.
- (3) Reliability of the entire test and four subtests was calculated with the Kuder-Richardson Formula #20.
- (4) Validity of the test was determined by a panel of science educators familiar with the content and process goals of the Science Curriculum Improvement Study, and by correlation with the Science Test (Form 4A) of the STEP (Series II) test.
- (5) The experimental design utilized in the study was the static-group comparison. The statistical test utilized was a repeated measures test with two treatment groups, SCIS and Control, and four measures: the SASAI, TSIS, STEP, and composite achievement scores on the State Assessment Examination.

CHAPTER V

SUMMARY AND CONCLUSIONS

Development of the Test

One major purpose of this study was to construct a reliable and valid paper and pencil group test for fifth grade students that would evaluate some of the stated or implied goals of grades 3, 4, and 5 of the Science Curriculum Improvement Study.

Items were written to satisfy the test specification grid, which included the main content areas specified in grades 3, 4, and 5 of the SCIS program, and four process areas of identifying and controlling variables, interpreting data, predicting, and inferring.

Validity for the test was provided by a panel of science educators, and correlations with Science Test (Form 4A) of the Sequential Test of Educational Progress (STEP) Series II. Reliability was determined by the Kuder-Richardson Formula #20.

Two pilot testing programs were held before the final form of the test was developed. The first tryout of the TSIS program involved 902 fifth grade students from the school districts of Van Dyke, Livonia, and East Lansing, Michigan. This first tryout was to clarify ambiguities in the instructions, format, and test items. Also, it provided data

for calculating test item difficulty and discrimination indices.

The second pilot test program included 185 fifth grade students from the school districts of Waterford, Van Dyke, and East Lansing, Michigan. The purpose of the second tryout program was to gather additional item analyses data. Many items on the revised preliminary form were altered on the basis of data gathered after the first pilot test. Item analyses data obtained from the second pilot test were the major determining factor for selecting items to be included in the final version of the TSIS.

The final form of the test, which contained 50 items, was administered to 310 fifth grade students in 12 classrooms in the school districts of DeWitt, Grand Ledge, and Perry, Michigan. Children in these schools had been in the SCIS program for five years. A Control group made up of 191 children in seven classrooms was selected from schools similar to the SCIS schools.

The final version of the TSIS was given to the above-mentioned sample of SCIS and Control students. Table 14 summarizes the significant data about the test.

Table 14.--TSIS analyses data.

	Index of Diffic.	Index of Discrim.	Stand. Error	Kuder-Richardson Formula #20
SCIS	46	49	3.0810	0.9033
Control	39	43	3.1017	0.8726

The Study

The second purpose of this study was to use the TSIS and other tests to compare fifth grade students who have been in the SCIS program for five consecutive years with similar fifth grade students who have been enrolled in traditional textbook series science courses.

Twelve SCIS and seven Control classrooms were compared using scores from the TSIS, SASAI, STEP (Series II, Form 4A Science Test), and the composite achievement score of the 1970-71 State of Michigan Assessment Examination. Results were analyzed using a repeated measures design, which indicated there was no main effect or significant difference between SCIS and Control classrooms on any of the four measures used in the study.

Limitations

The following limitations were recognized:

(1) There was no natural control group for this study because children were not randomly assigned to the type of science instruction they received. An attempt was made to match Control schools with the SCIS schools based on State Assessment scores. This procedure did locate schools that are similar to the experimental schools on measured parameters; however, they were not equal on all possible parameters. A more meaningful study could be conducted if students were assigned to science instruction randomly.

(2) The last tryout of Form B of the TSIS involved schools not included in the study, but fifth grade students who had five years of SCIS instruction in the proper sequence. The general achievement of these schools, as measured by the State Assessment, was considerably higher than the SCIS and Control schools used in the study. Since the major body of items for the final test were selected from Form B on the basis of their performance on this last tryout, perhaps items too difficult for the study sample were selected on the basis of this last tryout.

(3) The validity of the test may also be affected by the reading level and length of the test. The test is quite long (25 sides) and many fifth grade teachers felt that the reading level was difficult for their students.

Findings and Conclusions

The following hypotheses were tested:

H_{01} : There is no significant difference between classrooms receiving science instruction with SCIS and classrooms not receiving SCIS instruction as measured by the Science and Scientists Attitude Inventory (SASAI), the Test of Science Inquiry Skills (TSIS), the Sequential Test of Educational Progress (STEP), and the State of Michigan Assessment (composite achievement score).

When each classroom's combined means results for each test were compared, there was no significant difference between SCIS and Control classrooms. Referring to Figure 8

in Chapter IV, which depicts the mean scores of SCIS and Control classrooms, the graph shows that the Control schools had higher mean scores on the SASAI, TSIS, and STEP tests and that the SCIS schools were higher on the State Assessment. However, when all these scores were considered simultaneously, the difference between SCIS and Control classrooms was not significant.

The reason for this lack of main effect was that neither group had higher mean scores on all four tests. However, even if one group had scored higher on all four tests, the possibility still exists that the difference between groups on each test would not be great enough to produce a significant main effect.

As mentioned in Chapter III, State Assessment scores on socioeconomic status and composite achievement were utilized to identify Control schools that were very similar to the SCIS schools on overall measured parameters. However, the investigator felt that the children in the SCIS schools, because of the nature of the SCIS program philosophy and activities, would have more realistic attitudes toward science and scientists that could be measured by the Science and Scientists Attitude Inventory and that children in the SCIS schools would score higher on the Test of Science Inquiry Skills, which is based on SCIS objectives. It was further believed that the Sequential Test of Educational Progress, which contains predominantly recall-type questions, favored children in textbook series science programs. Essentially,

then, the investigator was, in fact, predicting no treatment main effect prior to the study. However, treatment by measures interaction was predicted, for reasons mentioned above.

Ho₂: On the Science and Scientists Attitude Inventory (SASAI), the Test of Science Inquiry Skills (TSIS), the Sequential Test of Educational Progress (STEP), and the State of Michigan Assessment, there will be no significant interaction between treatment and measures.

Analyses of the data indicate that when the difference between SCIS and Control classrooms was compared on each of the four tests, the treatment groups were not significantly different on any of the four tests.

Upon reviewing SCIS classroom mean scores on all four tests, a wide range of scores was noted; this fact was masked when only the composite SCIS classroom mean score was used. On the TSIS, for example, the composite score for all 12 SCIS classrooms of 49.31 gives no indication that the range of mean scores for these classrooms was from 42.27 to 53.54. Higher mean scores are cancelled out by the lower ones. The range of mean scores for the seven Control classrooms on the TSIS was from 47.18 to 54.77 and the composite score was 51.46. Thus, the composite score used in the analyses is not truly representative of the heterogeneous nature of the populations. For example, Classroom #8 of the SCIS population had the lowest scores on all four tests. A follow-up discussion with the principal of that school

indicated that the teacher of this classroom was an excellent disciplinarian and that he was given a disproportionate number of children who had been identified as discipline problems. In fact, based on their previous records, the principal was surprised that these students did as well as they had.

Another possibility is that with the use of only seven Control classrooms in the study, if one of these classrooms was considerably different from the SCIS classroom it was selected as a control for, this could have a significant weighting effect on the Control school scores used in the analyses.

Since there was no significant interaction, Hypotheses 3, 4, 5, and 6 were not considered.

Evidence from this study indicates that the paper and pencil test format can function effectively as a group test for fifth grade students, and that the test provides a means of evaluating a group of students' ability to utilize process skills. The panel of science educators who afforded validity to the test items indicated that the items included in the Test of Science Inquiry Skills do measure process skills. These items were then compiled into a single test which was administered simultaneously to classrooms of fifth grade students.

Feedback from teachers who administered the Test of Science Inquiry Skills indicated that the students had no difficulty in filling out and marking the machine scorable

answer sheets. Also, they had no difficulty with the test itself; they could follow the questions and relate the test items to specific diagrams. Diagrams, charts, graphs, and pictures can thus be used to effectively communicate problem situations to fifth graders.

The fact that SCIS students did as well as the Control students on the STEP test indicates that nothing, in the way of achievement, is lost by utilizing the SCIS program. None of the instruments used in the study were able to distinguish between the SCIS and Control populations. This may be because the students in the sample have reached equal achievement levels on all measured parameters. Also, other circumstances, as mentioned, may have been responsible for these results.

Implications for Education

As previously mentioned in Chapter III, the SCIS teachers whose classrooms were included in the study had not taught the SCIS program prior to the year that these children were in their classes. Thus, as the children progressed through the program each year, they were taught by a teacher who was teaching the SCIS program for the first time. It may well be that with such a short exposure to the SCIS program and its accompanying change in teaching style and emphasis that this program actually becomes very confusing to the teacher and subsequently to the student. Perhaps the SCIS population could not be distinguished from

the Control population for a reason such as this. It is predicted that these teachers will increase their effectiveness as SCIS teachers as they gain experience with the materials and philosophy of the SCIS program in succeeding years to come.

The maturation and experience of the students in the sample may also affect their attainment of process skills, regardless of what science program they have been exposed to. The students included in the study were all fifth grade students, and it is possible that children of this age level have already gained the intellectual skills for problem solving as measured in this study. Thus, it would be difficult to differentiate SCIS students from non-SCIS students at this age level.

Many research studies have been undertaken to determine differences between traditional and process-oriented science programs at the secondary level. While these have contributed greatly to the improvement of secondary science programs, the findings of such studies cannot be directly applied to elementary science education, one reason being that secondary science classrooms usually represent more selective populations than do elementary science classrooms. Many children choose not to take science at the high school level, whereas in the elementary school, all students are included in the science program. The lack of significant differences on all four instruments between fifth grade SCIS and Control populations could thus be explained by the fact

that at the elementary level, students have not been selected out, as they are in secondary science classes.

As discussed in Chapter IV, the STEP and TSIS tests were highly correlated with reading scores. If the reading factor of the TSIS could be reduced or made negligible, perhaps one could assess the problem-solving achievement skills of a student more accurately.

The TSIS assesses problem solving in four areas: identifying and controlling variables, interpreting data, predicting, and inferring. These process areas were dealt with as whole units, and did not take into consideration that various types of problem solving require different subprocess skills. Thus, if a student failed to solve a problem within a specified process area on the TSIS, one could not determine at what level in the spectrum of lower level to more complicated problem-solving skills the student failed.

Another question arises regarding the relationship between a student's understanding of SCIS concepts and his performance on the TSIS. Theoretically, the concepts that a student brings to a problem-solving or process test should not affect his performance on that test.

Suggestions for Future Research

Based on the realization that all SCIS teachers whose classrooms were utilized in the study had not taught the SCIS program prior to the year that these students

entered their classrooms, a longitudinal study of these SCIS teachers two or three years hence may prove valuable. Does a teacher, and consequently her students, take time to adjust to a new program? Are initial setbacks overcome eventually by positive gains once a teacher becomes more familiar with the materials and philosophy of a new program? These and other questions need to be answered, not only as it concerns this study, but also all new science programs as well.

Another suggestion for future research would be to utilize individual competency measures, such as the AAAS Competency Measures, to compare rural with urban students, and students with and without the new process-centered programs. This would shed some light on what effect a child's background, personal experiences, and maturity have on his ability to perform intellectual skills such as problem solving.

The TSIS, as developed for this study, could be used in a study using multiple factor analyses with a similar population. This type of study would require a maximum amount of information regarding each student in the experimental and control populations, yet it may reveal in more precise terms the factors which affect the students' problem-solving test scores.

As mentioned earlier, the TSIS does not determine the level of subprocess skills that a student has attained. It is here suggested that a problem-solving test be developed

that would measure various levels in the process skills hierarchy. This would not only reveal more information about the student, but could also be used as a diagnostic tool to further guide the student in his learning experience.

Perhaps the TSIS could be put into a format that does not require any reading. For example, taped oral questions and instructions could be synchronized with slides showing the diagrams and pictures.

A companion test measuring primarily SCIS content areas could be developed. It would then be possible to determine the relationship between understanding of SCIS concepts and problem-solving ability.

BIBLIOGRAPHY

BIBLIOGRAPHY

Books

- American Association for the Advancement of Science. Science--A Process Approach: An Evaluation Model and Its Application. Second report. Washington, D. C.: American Association for the Advancement of Science, Misc. Pub. 68-4, 1968.
- Biological Sciences Curriculum Study. Resource Book of Test Items for Biological Science: Molecules to Man. Revised ed. Educational Programs Improvement Corp., 1971.
- Bloom, Benjamin S., et al. Taxonomy of Educational Objectives Handbook I: Cognitive Domain. New York: David McKay Company, Inc., 1956.
- _____; Hastings, J. Thomas; and Madaus, George F. Handbook on Formative and Summative Evaluation of Student Learning. New York: McGraw-Hill, 1971.
- Butts, David P. An Inventory of Science Methods. Austin, Texas: Science Education Center, University of Texas, 1966.
- Campbell, Donald T., and Stanley, Julian C. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally & Co., 1963.
- Conrad, Herbert S. "The Experimental Tryout of Test Materials." Educational Measurement. Washington, D. C.: American Council on Education, 1951.
- Cooperative Tests and Services. Bulletin of Tests and Services. Princeton, New Jersey: Educational Testing Service, 1971.
- _____. SCAT-STEP Series II Book of Norms. Princeton, New Jersey: Educational Testing Service, 1970.
- Ebel, Robert L. Essentials of Educational Measurement. 2nd ed. Englewood Cliffs, New Jersey: Prentice Hall, 1972.

Final Report on SCIS Activities for the 1968-69 School Year
at the Magnolia Public Schools. Magnolia, Arkansas,
1969.

Flory, Donald L. "Dynamics of Classroom Behavior: An Informal Study of SCIS." What Is Curriculum Evaluation? Edited by Robert Karplus. Berkeley: University of California, 1968.

Gagné, Robert M. "Psychological Issues in Science--A Process Approach." The Psychological Bases of Science--A Process Approach. Washington, D. C.: American Association for the Advancement of Science, Misc. Pub. 65-8, 1965.

Gronlund, Norman E. Measurement and Evaluation in Teaching. 2nd ed. New York: The Macmillan Company, 1971.

Hunt, Joseph M. Intelligence and Experience. New York: The Ronald Press Company, 1961.

Hurd, Paul D. "Elementary Science for the 1970's in Perspective." Developments in Elementary School Science. Washington, D. C.: American Association for the Advancement of Science, Misc. Pub. 70-18, 1970.

Ipsen, D. C. Issues in Elementary Science. Washington, D. C.: National Science Teachers Association, 1970.

Karplus, Robert, and Thier, Herbert D. A New Look at Elementary Science. Chicago: Rand McNally & Company, 1967.

Kessen, William. "Statement of Purposes and Objectives of Science Education in School." The Psychological Bases of Science--A Process Approach. Washington, D. C.: American Association for the Advancement of Science, Misc. Pub. 65-8, 1965.

Kirk, Roger E. Experimental Design: Procedures for the Behavioral Sciences. Belmont, California: Brooks/Cole Publishing Co., 1968.

Klopfer, Leopold E. "Evaluation of Learning in Science." Handbook on Formative and Summative Evaluation of Student Learning. New York: McGraw-Hill, 1971.

Lockard, J. D., ed. Seventh Report of the International Clearinghouse on Science and Mathematics Curriculum Improvements. College Park, Maryland: University of Maryland, 1970.

Mehrens, William A., and Lehmann, Irvin J. Standardized Tests in Education. New York: Holt, Rinehart & Winston, Inc., 1969.

Michigan Department of Education. 1970-71 Individual Pupil Report: Explanatory Materials. Michigan Educational Assessment Program, April, 1971.

_____. Local District Report: Explanatory Materials. Michigan Educational Assessment Program, Report No. 6, 1970.

Nedelsky, Leo. Science Teaching and Testing. New York: Harcourt, Brace & World, Inc., 1965.

Nelson, Clarence H. Measurement and Evaluation in the Classroom. Toronto: Macmillan Company, 1970.

Ness, Patricia D. "Organisms Feedback." What Is Curriculum Evaluation? Edited by Robert Karplus. Berkeley: University of California, 1968.

Rimoldi, H. J. A. Training in Problem Solving. Chicago: Loyola University Press, 1961.

Shulman, Lee S. "Psychology and Mathematics Education." The Sixty-Ninth Yearbook of the National Society for the Study of Education, Part I. Edited by Edward G. Begle. Chicago: University of Chicago Press, 1970.

Science Curriculum Improvement Study. Interaction and Systems: Evaluation Supplement. Berkeley: University of California, June, 1971.

_____. Life Cycles: Evaluation Supplement. Berkeley: University of California, June, 1971.

Siegelman, Ellen, and Karplus, Elizabeth. "Relativity Feedback Testing." What Is Curriculum Evaluation? Edited by Robert Karplus. Berkeley: University of California, 1968.

Suchman, J. R. The Elementary School Training Program in Scientific Inquiry. Urbana: University of Illinois, June, 1962.

Tresmontan, Olympia D. "SCIS in the Classroom, A Sample." What Is Curriculum Evaluation? Edited by Robert Karplus. Berkeley: University of California, 1968.

Periodicals

- Allen, Leslie R. "An Examination of the Ability of First Graders From the SCIS Program to Describe an Object by Its Properties." Science Education, LV, 1 (1971), 60-67.
- Anderson, Ronald D. "Evaluation in Elementary School Science, Part I: Formulating Objectives." Science and Children, V (September, 1967), 20-23.
- Atkin, J. M. "Science in the Elementary School." Review of Educational Research, XXXIV, 3 (1964), 267.
- _____. "Some Evaluation Problems in a Course Content Improvement Project." Journal of Research in Science Teaching, I, 1 (1963), 129-32.
- Austin, Gilbert R. "Educational Accountability--Hallmark of the 1970's." The Science Teacher, XXXVIII (April, 1971), 26-28.
- Beard, Jean. "The Development of Group Achievement Tests for Two Basic Processes of AAAS Science--A Process Approach." Journal of Research in Science Teaching, VIII, 2 (1971), 179-83.
- Butts, David P. "The Relationship of Problem-Solving Ability and Science Knowledge." Science Education, XLIX, 2 (1965), 138-46.
- _____, and Jones, Howard L. "Inquiry Training and Problem-Solving in Elementary School Children." Journal of Research in Science Teaching, IV, 1 (1966), 21-27.
- _____. "The Development of the TAB Science Test." Science Education, LI (December, 1967), 463-73.
- Cross, K. P., and Gaier, Eugene L. "Technique in Problem Solving as a Predictor of Educational Achievement." Journal of Educational Psychology, XLVI, 4 (1955), 193-206.
- Das, Radha C. "The Problem Development Method." Science Education, XLVIII, 5 (1964), 485-88.
- Dietz, Maureen A., and George, Kenneth D. "A Test to Measure Problem Solving Skills in Science of Children in Grades 1, 2, and 3." Journal of Research in Science Teaching, VII, 4 (1970), 341-51.

- Dressel, P. L. "Better Tests for Science Classrooms." The Science Teacher, XXVI (September, 1959), 342-45.
- Fischler, Abraham S. "Change in Classroom Behavior Resulting From an In-Service Program Utilizing Television." School Science and Mathematics, LXVII (April, 1967), 321-24.
- _____. "Implications of Structure for Elementary Science." Science Education, LII, 3 (1968), 278-79.
- Gagné, Robert M. "Why the 'Process' Approach for a Modern Curriculum?" Educational Products Information Exchange, I, 8 (1968), 11-12.
- Glass, Lynn W. "Assessment of Affective Outcomes of Instruction With High School Sophomore Biology Students and Teachers." Journal of Research in Science Teaching, VII, 1 (1971), 41-43.
- Howe, Ann C., and Butts, David P. "The Effect of Instruction on the Acquisition of Conservation of Volume." Journal of Research in Science Teaching, VII, 4 (1970), 371-75.
- Klopfer, Leopold E. "Effectiveness and Effects of ESSP Astronomy Materials--An Illustrative Study of Evaluation in a Curriculum Development Project." Journal of Research in Science Teaching, VI, 1 (1969), 64-75.
- Lyda, Wesley J., and Morse, Evelyn C. "Attitudes, Teaching Methods, and Arithmetic Achievement." Arithmetic Teacher, VIII (March, 1961), 117-19.
- Mallinson, Jacqueline B. "The Current Status of Science Education in the Elementary School." School Science and Mathematics, LXI (April, 1961), 252-70.
- Mayer, William V. "Evaluation and Curriculum Development." BSCS Newsletter, 46 (February, 1972), 1-2.
- Mead, Margaret, and Métraux, Rhoda. "Image of the Scientist Among High School Students." Science, CXXVI (August, 1957), 384-90.
- Morgan, David A. "STEPS--A Science Test for Evaluation of Process Skills." The Science Teacher, XXXVIII (November, 1971), 77-79.
- National Science Teachers Association. "School Science Education for the 1970's." The Science Teacher, XXXVIII (November, 1971), 45-51.

- Nay, Marshall A., and Associates. "A Process Approach to Teaching Science." Science Education, LV, 2 (1971), 197-207.
- Nicodemus, Robert B. "Content and Skill Hierarchies in Elementary Science: An Analysis of ESS Small Things." Journal of Research in Science Teaching, VII, 3 (1970), 173-77.
- Ramsey, Gregor A., and Howe, Robert W. "An Analysis of Research Related to Instructional Procedures in Elementary School Science." Science and Children, VI (April, 1969), 25-36.
- Raun, Chester E., and Butts, David P. "The Relationship Between the Strategies of Inquiry in Science and Student Cognitive and Affective Behavior Change." Journal of Research in Science Teaching, V, 3 (1967-68), 261-68.
- Science Curriculum Improvement Study. SCIS Newsletter. Edited by Suzanne Stewart. No. 17 (1970).
- Tannenbaum, Robert S. "The Development of the Test of Science Processes." Journal of Research in Science Teaching, VII, 2 (1971), 123-36.
- Taylor, Alton L. "Curriculum and Instructional Evaluation in Science." Science Education, LIV, 3 (1970), 237-39.
- Thier, Herbert D. "A Look at First Graders' Understanding of Matter." Journal of Research in Science Teaching, III (March, 1965), 84-89.
- Thompson, Barbara S., and Voelker, Alan M. "Programs for Improving Science Instruction in the Elementary School, Part II: SCIS." Science and Children, VII, 8 (1970), 29-37.
- Walbesser, Henry H. "Science Curriculum Evaluation: Observations on a Position." The Science Teacher, XXXIII (February, 1966), 34-39.
- _____, and Carter, Heather L. "Some Methodological Considerations of Curriculum Evaluation Research." Educational Leadership, XXVI (October, 1968), 53-64.
- _____. "The Effect on Test Results of Changes in Task and Response Format Required by Altering the Test Administration From an Individual to a Group Form." Journal of Research in Science Teaching, VII, 1 (1970), 1-8.

Welch, Wayne W. "Curriculum Evaluation." Review of Educational Research, XXXIX, 4 (1969), 429-43.

_____. "The Need for Evaluating National Curriculum Projects." Phi Delta Kappan, LXIX, 9 (1968), 530-32.

_____, and Pella, Milton O. "The Development of an Instrument for Inventorying Knowledge of the Processes of Science." Journal of Research in Science Teaching, V, 1 (1967), 64-68.

Unpublished Works

Battaglini, Dennis W. "An Experimental Study of the Science Curriculum Improvement Study Involving Fourth Graders' Ability to Understand Concepts of Relative Position and Motion Using the Planetarium as a Testing Device." Unpublished Ph.D. dissertation, Michigan State University, 1971.

Benson, Keith S. "A Comparison of Two Methods of Teaching Fifth Grade Science." Unpublished Ph.D. dissertation, Oklahoma State University, 1968.

Edwards, Thomas F. "A Study in Teaching Selected Intellectual Skills to Kindergarten Children." Unpublished Ed.D. thesis, Michigan State University, 1966.

Fyffe, Darrel W. "The Development of Test Items for the Integrated Science Processes: Formulating Hypotheses and Defining Operationally." Unpublished Ph.D. dissertation, Michigan State University, 1971.

Geesaman, Marjorie W. "Evaluating the Educational Outcomes of Process Education in Science by Objective Test." Paper presented at the 20th meeting of the National Science Teachers Association, New York, April, 1972.

Motz, LaMoine L. "The Development of an Instrument to Evaluate Sixth and Ninth Grade Students' Attitudes Towards Science and Scientists." Unpublished Ph.D. dissertation, University of Michigan, 1970.

Partin, Melba S. "An Investigation of the Effectiveness of the AAAS Process Method Upon the Achievement and Interest in Science for Selected Fourth Grade Students." Unpublished Ph.D. dissertation, University of Southern Mississippi, 1967.

- Robison, Richard Wayne. "The Development of Items Which Assess the Processes of Controlling Variables and Interpreting Data." Unpublished Ph.D. dissertation in process, Michigan State University, 1970.
- Stoker, Howard W. "Aptitude and Attitudes of High School Youth in Regard to Science as Related to N Variables." Unpublished Ph.D. dissertation, Purdue University, 1957.
- Welch, Wayne W. "The Development of an Instrument for Inventorying Knowledge of the Processes of Science." Unpublished Ph.D. dissertation, University of Wisconsin, 1966.

APPENDICES

APPENDIX A

A TEST OF SCIENCE INQUIRY SKILLS
AND SAMPLE ANSWER SHEET

A TEST OF SCIENCE INQUIRY SKILLS

by Joseph W. Riley

DIRECTIONS

1. Remove the answer sheet from the test booklet. Close the booklet after removing the answer sheet.
2. Print your name, last name first, in the space provided in the upper left corner of the answer sheet.
3. In the space next to your name, write your teacher's name.
4. In the spaces under your name, write your birth date, for example, January 4, 1961.
5. In the boxes in the upper right corner print in as much of your name as possible and fill in the correct spaces under your name.
6. Find the box labeled "sex" at the bottom of the page. If you are a boy, fill in the space labeled "M". If you are a girl, fill in the space labeled "F".
7. Now turn the answer sheet so that your name is on the left side of the page.
8. Find Number 1 on the answer sheet, then find Number 8. Notice that the numbers run ACROSS the page, from left to right.
9. Now open your test booklet. Each question in this test is followed by 3 or 4 choices for an answer. Read each question and then decide which is the one best answer.
10. On the answer sheet find the row which has the same number as the question you are answering. In that row blacken the space that has the number which is the same as the one in front of the answer you choose.
11. Make your marks heavy and black, press down hard on your pencil, and go over your marks. If you make a mistake, erase completely the answer you wish to change. Use only the special pencil that has been provided.
12. Look at this sample question:

85. A cow is a(n):

- | | |
|------------|-----------|
| 1. machine | 3. animal |
| 2. plant | 4. rock |

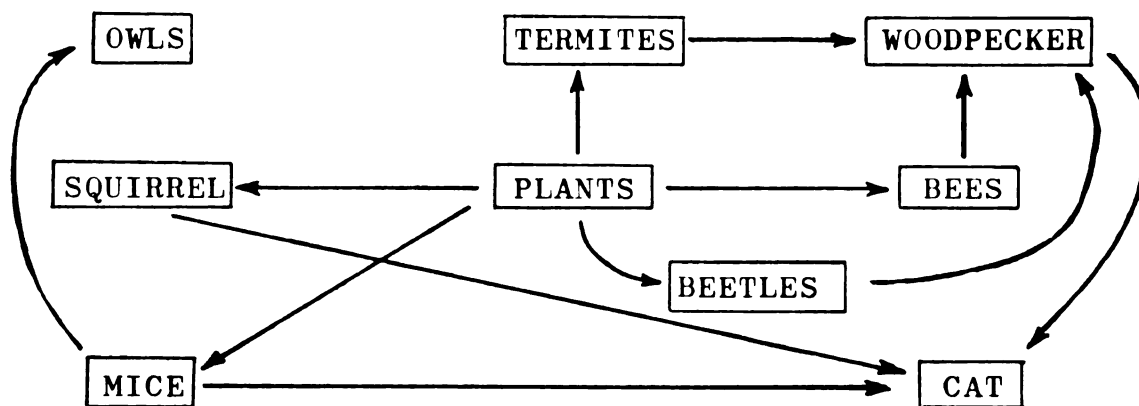
For this question, the answer sheet should look like this:

85. ☐ 1 ☐ 2 ☒ 3 ☐ 4

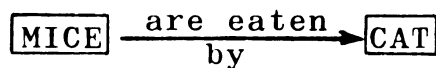
For each item in this examination select the ONE best answer. Do not mark more than ONE answer space for each item.

* * *

For items 1-6 use the following picture and example:



KEY: For example --



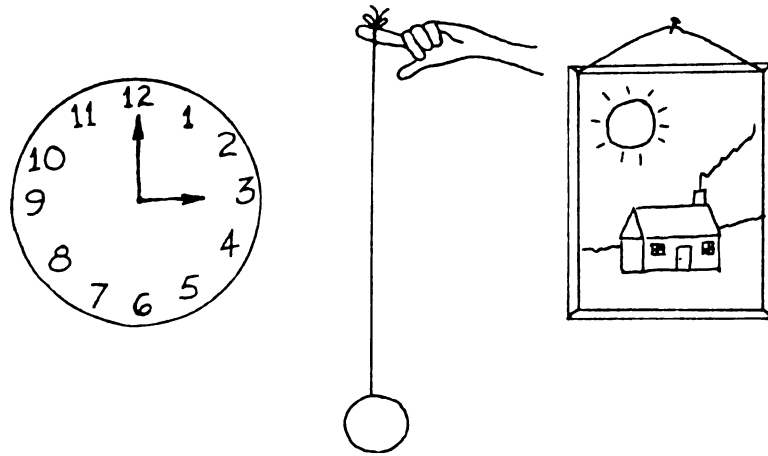
1. This food web shows that
 1. Owls eat cats
 2. Mice eat plants, and owls and cats eat mice
 3. Bees eat beetles
 4. Woodpeckers eat termites, bees, and mice
2. If the owls were removed from this area, one would expect an increase in the number of
 1. bees 2. plants 3. woodpeckers 4. mice
3. If the owls stayed away forever, probably the number of _____ would increase.
 1. bees 2. cats 3. woodpeckers 4. beetles
4. If more woodpeckers were to move into the area we would expect
 1. an increase in the number of termites
 2. a decrease in the number of beetles
 3. an increase in the number of squirrels
 4. a decrease in the number of plants

5. If cats are removed from this area, the number of
 1. owls would decrease
 2. woodpeckers would decrease
 3. mice would increase
 4. plants would increase

6. The area was sprayed with DDT (a poison) to kill the beetles. A number of owls were found dead due to DDT.
NOTE: These owls don't eat beetles. We can conclude that the owls
 1. got the DDT from the mice who got it from the plants
 2. got the DDT from the beetles
 3. got the DDT from the plants
 4. were not affected by the DDT

* * * *

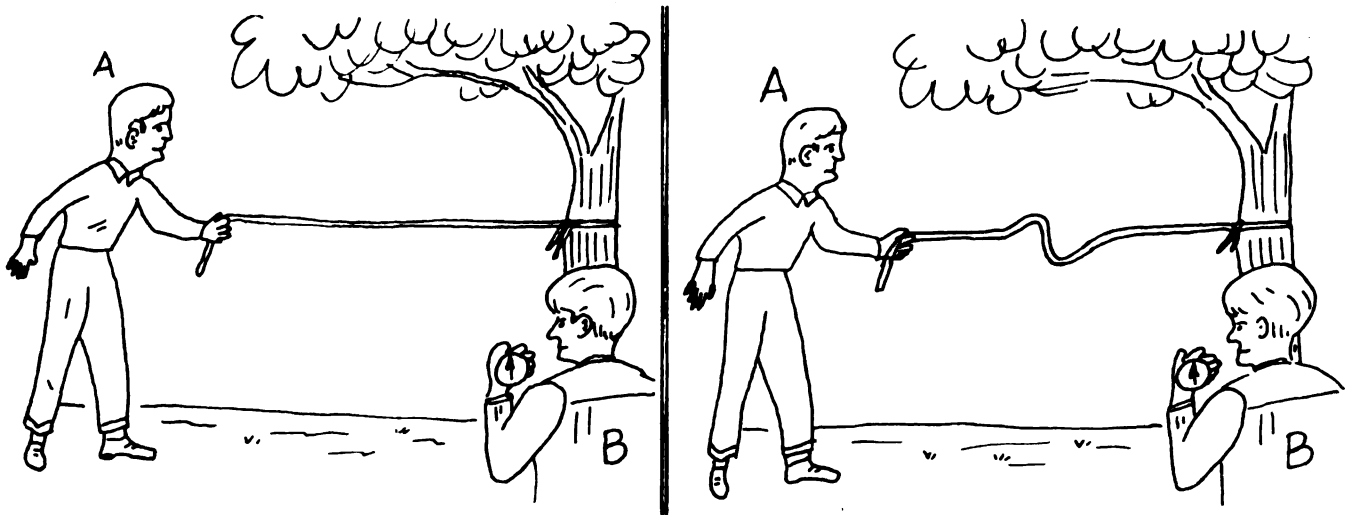
Item 7 refers to the following diagram:



7. Suppose you are trying to find out how long it takes the ball to swing back and forth. What variable in the system would make a difference in the time?
1. The strength of the person who is holding the string.
 2. The length of the string
 3. The surface of the ball (whether it is rough or smooth)
 4. The time on the clock

* * * *

Items 8 & 9 refer to the following pictures:



8. In the pictures above, what is the energy source that caused the change in the rope's pattern?
- | | |
|--------------|-------------|
| 1. The tree | 3. The rope |
| 2. The clock | 4. Boy A |
9. If more energy were used in moving the end of the rope up and down, how would the pictures be different?
1. Boy A would be taller
 2. The clock would run faster
 3. The size of the loop would be larger
 4. The rope would be longer

* * * * *

Items 10-11 refer to the following experiment:

Three mystery powders labeled A, B, and C are tested.

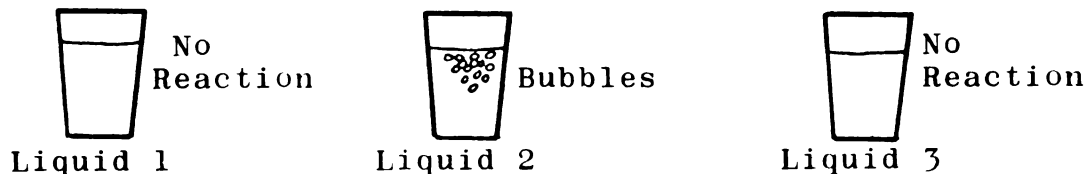
STEP I: A pinch of Powder A is put into each of three jars containing different colorless liquids and labeled Liquid 1, 2, and 3. (see diagram)

STEP II: Next, a pinch of Powder B is put into a second set of jars labeled Liquid 1, 2, and 3.

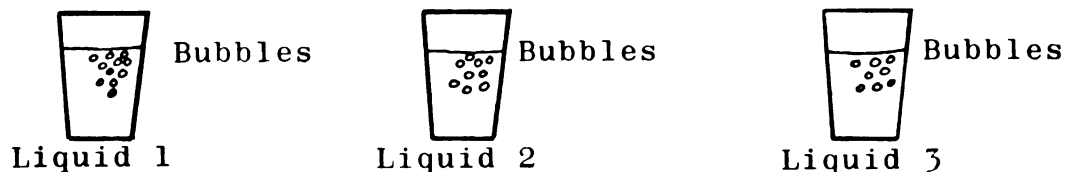
STEP III: Finally, small amounts of Powder C are put into a third set of jars labeled Liquid 1, 2, and 3. The results are shown in the diagram below.

DIAGRAM

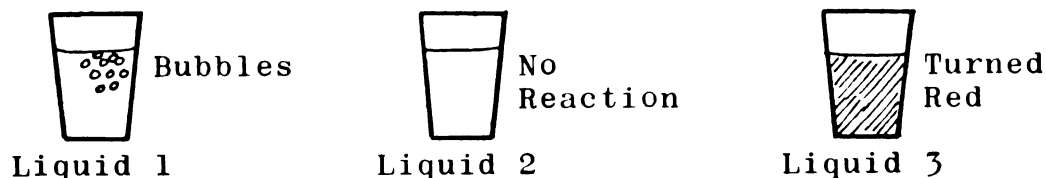
POWDER A:



POWDER B:



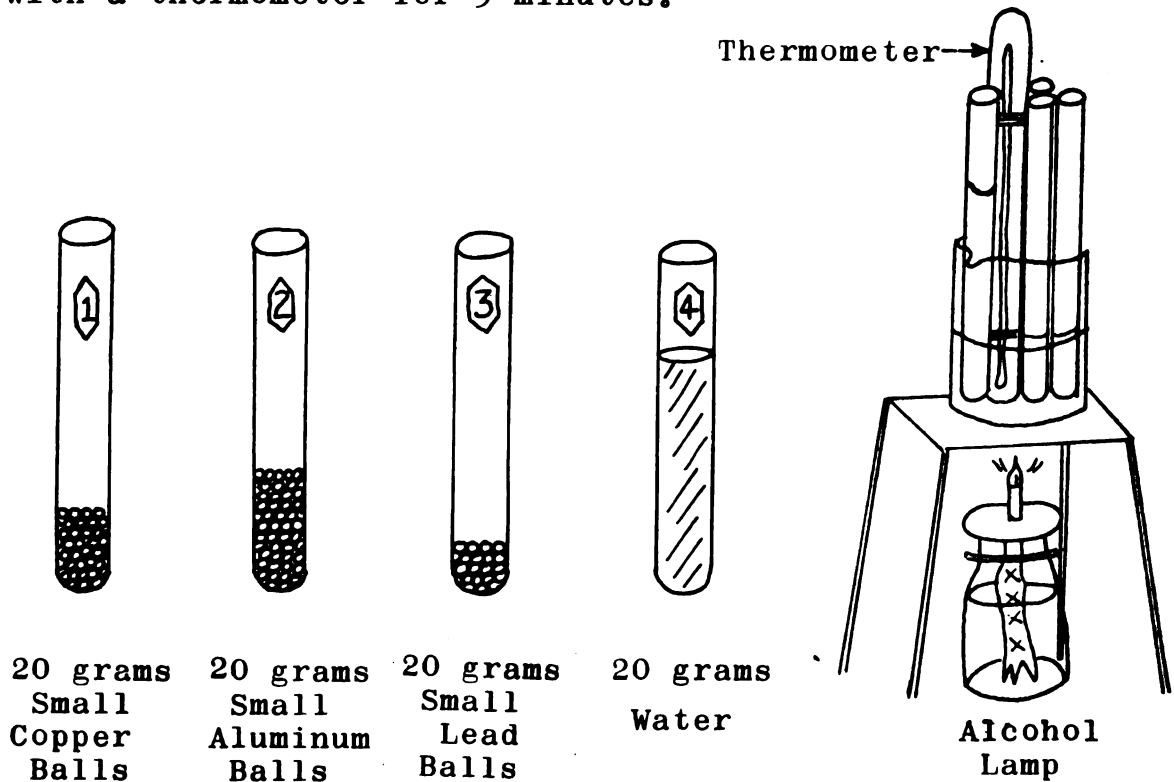
POWDER C:



10. Suppose Powder C was added to Mystery Liquid Q producing a red color. Mystery Liquid Q could be
1. Liquid 1 or 3
 2. Liquid 1, 2, or 3
 3. Liquid 3 only
 4. Liquid 2 only
11. Suppose Powder B and Mystery Liquid R produced bubbles. We could conclude that Mystery Liquid R could be
1. Liquid 1, 2, or 3
 2. Liquid 1 or 2 only
 3. Liquid 2 only
 4. Liquid 3 only
12. Suppose a mystery powder was added to each of the liquids and the following results were produced:
- Mystery Powder X + Liquid 1 produced bubbles
Mystery Powder X + Liquid 2 produced no reaction
Mystery Powder X + Liquid 3 produced a red color
- From these results, the mystery powder X is probably:
1. Powder A
 2. Powder B
 3. Powder C
 4. None of these

* * * * *

Items 13-16 refer to the following experiment which tries to find out which material is the best heat source. The four test tubes were placed in a jar of boiling water with a thermometer for 5 minutes.



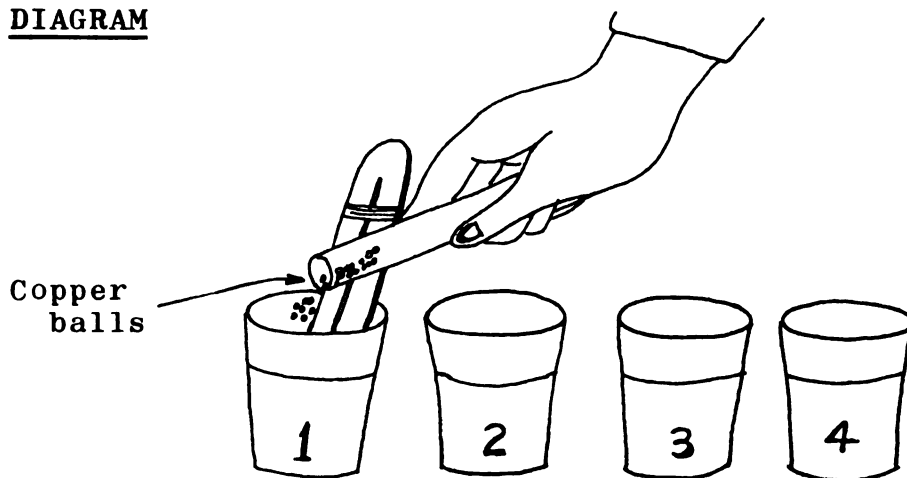
13. Which of these are variables in this experiment?

- | | |
|-----------------------|------------------------|
| 1. temperature | 3. weight of the balls |
| 2. kind of balls used | 4. time over the lamp |

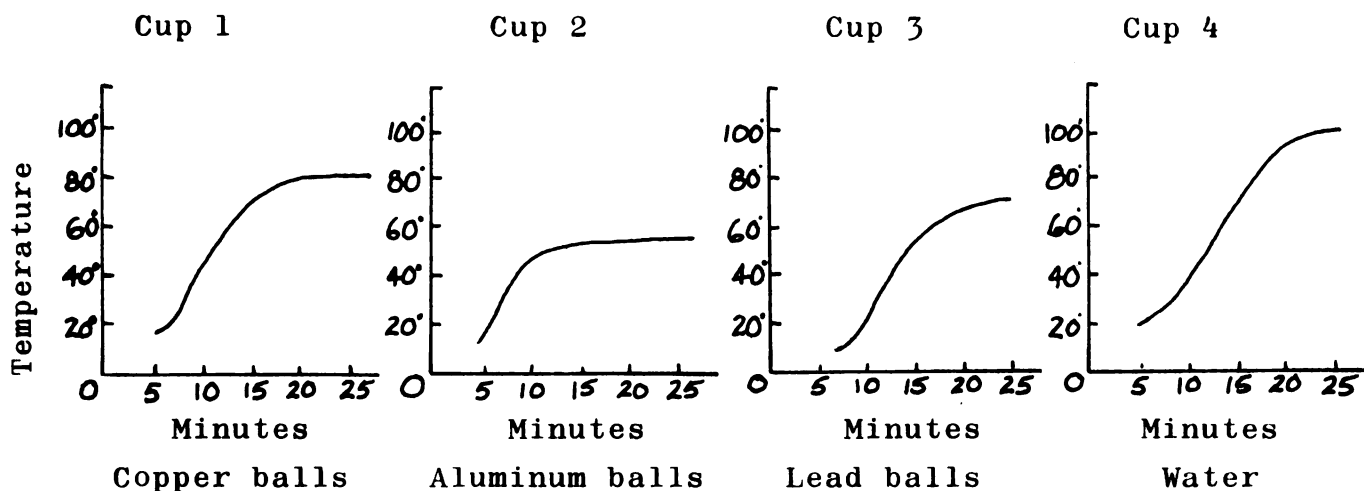
14. What is the immediate energy source?

- | | |
|----------------------|--------------------|
| 1. the boiling water | 3. the jar |
| 2. the test tubes | 4. the thermometer |

After the test tubes had been heated for 5 minutes, they were removed from the jar and their contents poured into separate cups, each of which was half full of water.
(see diagram on opposite page)

DIAGRAM

The graphs below shows the following results of the 4 cups:



15. Which cup shows the greatest increase in temperature?

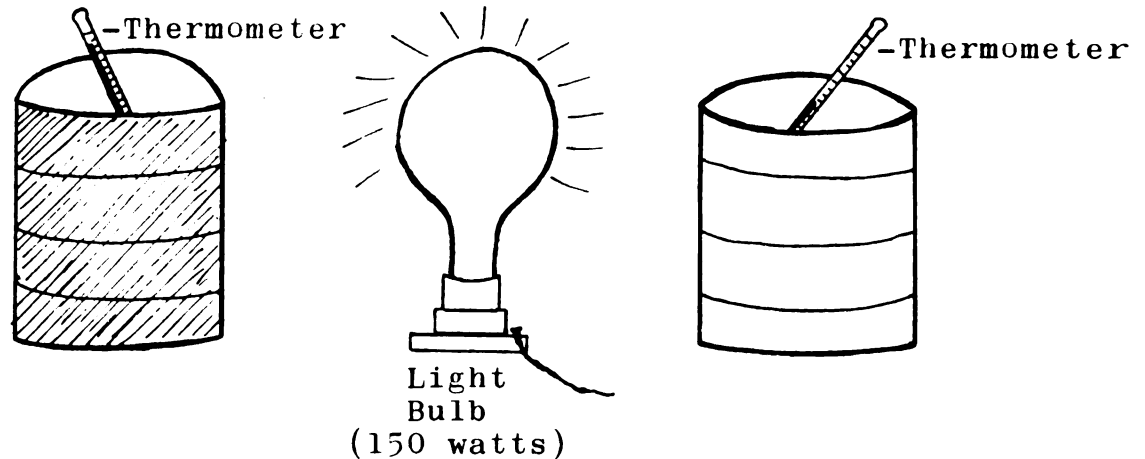
- | | |
|----------|----------|
| 1. Cup 1 | 3. Cup 3 |
| 2. Cup 2 | 4. Cup 4 |

16. Which material was the best heat source?

- | | |
|-------------------|---------------|
| 1. Copper balls | 3. Lead balls |
| 2. Aluminum balls | 4. Water |

* * * * *

Items 17-20 are based on the following drawing & explanation:



In this experiment we have a light bulb and 2 cans that are the same size. Each can is filled with the same amount of water. The outside of one can is painted black and the outside of the other is painted white. These cans are placed at equal distances from a light bulb that is turned on. After 3 hours, the water in the black can is hotter than the water in the white can.

17. The best energy receiver would be
 1. the light bulb
 2. the black can
 3. the white can
 4. both cans are equal

18. The energy source in this experiment would be the
 1. light bulb
 2. black can
 3. white can
 4. thermometer

19. The variable being tested in this experiment is the
1. amount of water used
 2. size of the can
 3. color of the can
 4. distance between the light and the cans
20. During the daytime in the summer, a house with a white roof should be
1. warmer than a house with a black roof
 2. cooler than a house with a black roof
 3. the same temperature as a house with a black roof

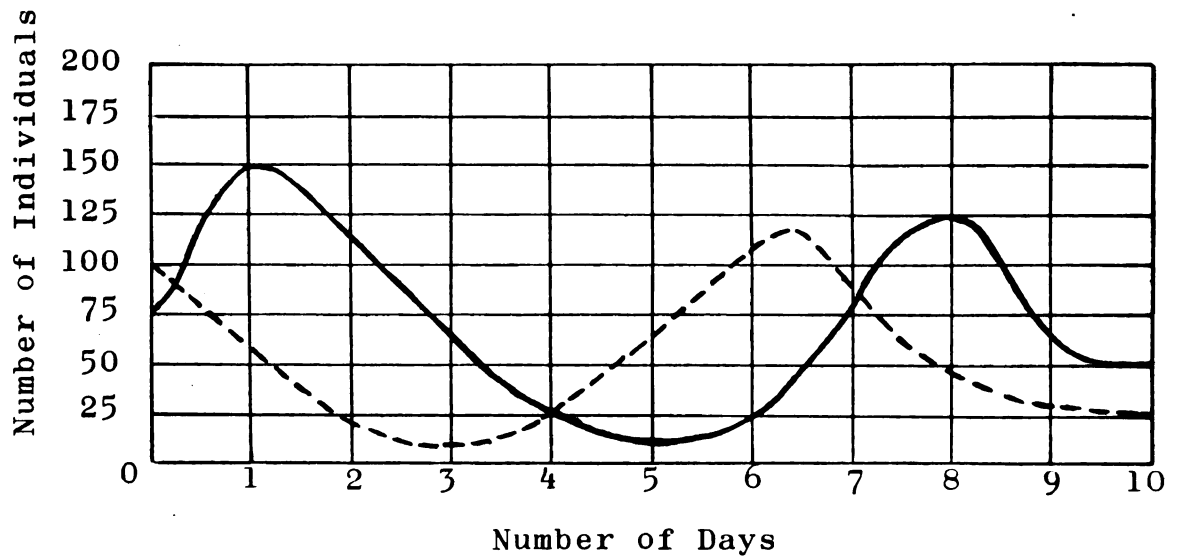
* * * * *

Items 21-24 refer to the following explanation and graph:

The following questions deal with a community made up of 2 different kinds of water animals. All the animals are living in the same small aquarium. The graph below describes the activities of these two populations.

———— = Population A

- - - - - = Population B



21. At the 8th day, what is the number of individuals in Population A?
1. 50
 2. 75
 3. 100
 4. 125

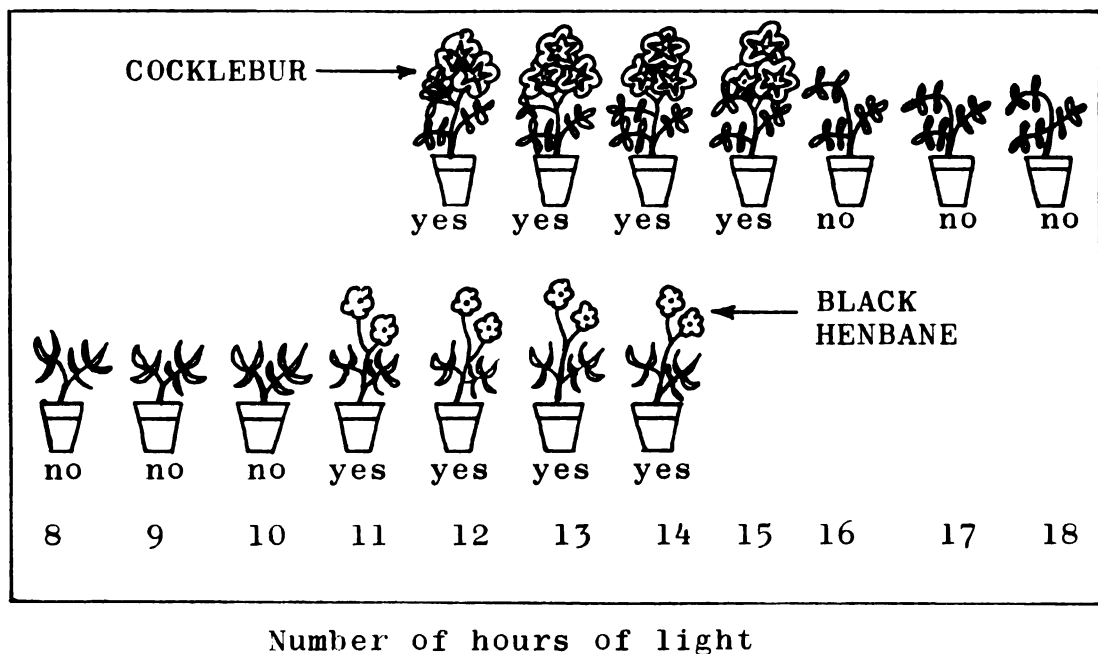
22. How many days are included on this graph?
1. 5
 2. 10
 3. 100
 4. 200
23. According to the graph, on which of these days was the number of individuals in each of the 2 populations exactly the same?
1. 2nd day
 2. 4th day
 3. 5th day
 4. 10th day
24. Why did the number of A individuals decrease after the 8th day?
1. They ran out of food to eat
 2. They had more babies
 3. They were eaten by the snails

* * * * *

Items 25-30 refer to the following explanation and tables:

Seven Cocklebur plants were used in an experiment.

(Cocklebur is a common plant that has pretty flowers.) The experiment was started when all of the plants were $2\frac{1}{2}$ inches tall. Each plant had a "light tight" cover that could be placed over the plant to keep out all light. Each plant was taken out of its box and exposed to a certain amount of light each day. For example, one plant was exposed to only 12 hours of light each day, another plant was exposed to 13 hours of light each day, and so on. Read the table carefully.



25. What variable is being investigated in this experiment?

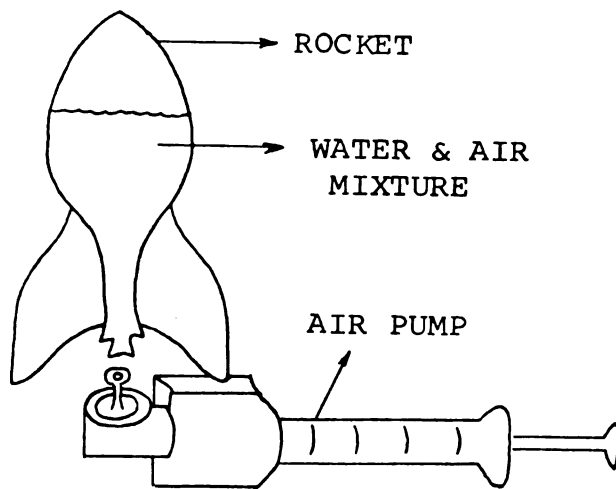
1. The amount of light plants receive
2. The height of the plants
3. The color of the flowers
4. The number of flowers on each plant

26. What conclusion can you draw from this experiment?
1. Cocklebur plants in light produce red flowers
 2. The amount of light a cocklebur plant receives may determine whether or not it will produce flowers
 3. Plastic covers cause cocklebur plants to produce flowers
 4. Cocklebur plants must be 3 inches tall to produce flowers

A second experiment was done using a different kind of plant. (Black Henbane plants) Some of these plants were exposed to light for 8, 9, 10, 11, 12, 13 or 14 hours each day. The students then looked at the plants to see which ones had flowers. Read the table carefully.

27. How many hours of light do these plants need in order to produce flowers?
1. 8 hours
 2. 10 hours
 3. 12 hours
 4. This cannot be determined
28. If you were to expose a Black Henbane plant to 6 hours of light, what do you predict would happen? The plant would:
1. die
 2. need water
 3. have flowers
 4. not produce flowers
29. If you wanted to grow both Cocklebur and Black Henbane plants together and have BOTH of them produce flowers, how many hours of light would you expose them to?
1. 8 to 10 hours
 2. 12 to 14 hours
 3. 16 to 18 hours
 4. not possible to decide
30. If you wanted to grow these plants together and have NEITHER of them produce flowers, how many hours of light would you expose them to?
1. 8 to 10 hours
 2. 12 to 14 hours
 3. 16 to 18 hours
 4. The experiment doesn't provide enough information to answer this.

Items 31 and 32 refer to the following diagrams & explanations:



PARTS:

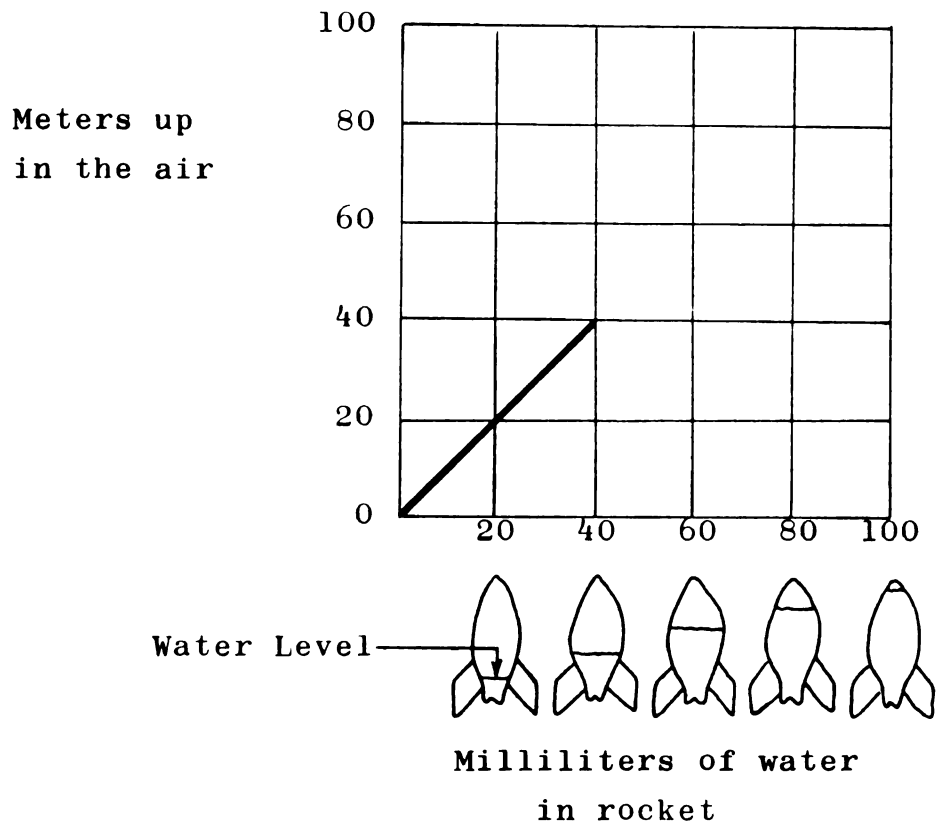
I. Air Pump

II. Rocket

III. Water & Air
Mixture

Perhaps you have seen a water rocket. It consists of the rocket itself and an air pump which fits on the bottom of the rocket. The person using the rocket adds the amount of water he wishes to the rocket, then places the air pump on the bottom of the rocket and pumps air into the rocket. The amount of pressure increases as the number of pumps is increased.

Suppose you were given the following information: Some tests were made using rockets with different amounts of water as shown on the bottom of the graph (next page) and how high the rocket rose on the side of the graph. Each rocket was given the same number of pumps of air. Study the graph on the next page carefully. You may write on the graph if you wish.



31. How high would you predict the rocket with 60 ml of water would go?

- | | |
|--------------|---------------|
| 1. 45 meters | 3. 80 meters |
| 2. 60 meters | 4. 100 meters |

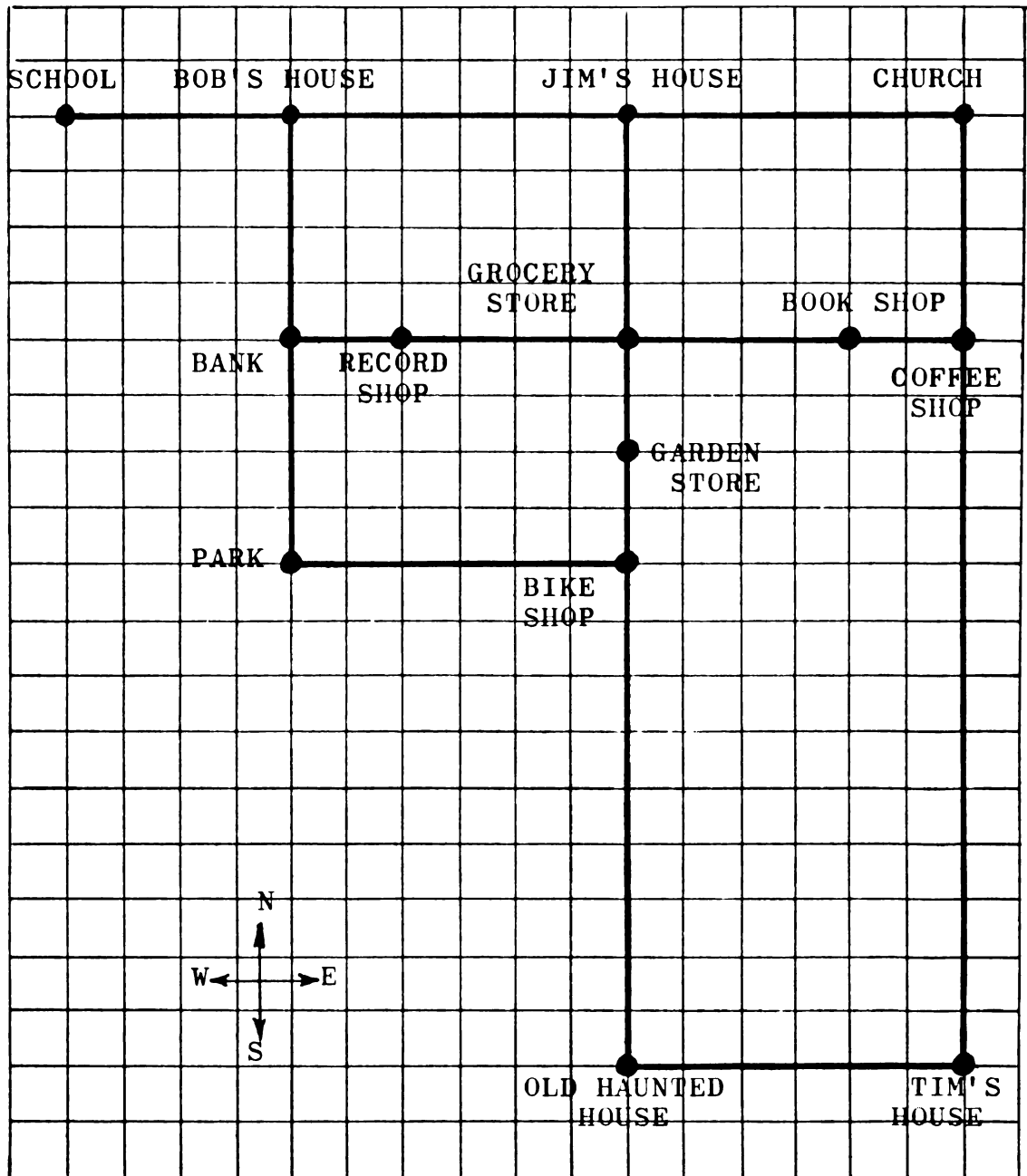
When a rocket with 70 ml of water in it was fired, it was found to go up 45 meters. When the rocket with 80 ml of water in it was fired, it was found to go up 40 meters.

32. Knowing this, how high would you predict the rocket with 100 ml of water in it would go?

- | | |
|--------------|--------------|
| 1. 35 meters | 3. 60 meters |
| 2. 40 meters | 4. 80 meters |

* * * * *

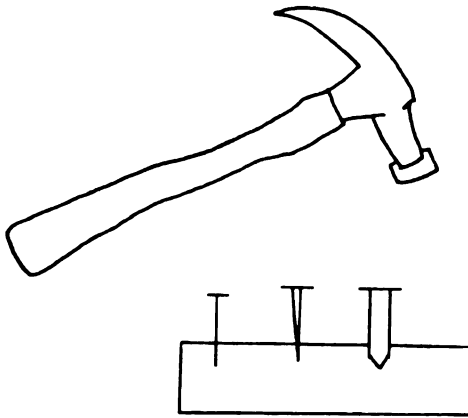
Items 33-37 refer to the following map. On this map, each space between 2 lines represents a distance of 1 unit. The compass in the bottom left corner tells the directions.



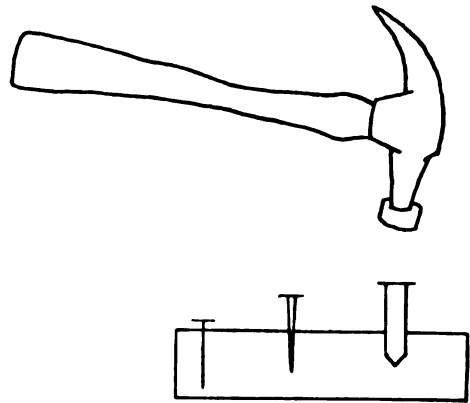
33. How many units is it from the school to Jim's house?
- | | |
|------------|-------------|
| 1. 6 units | 3. 10 units |
| 2. 8 units | 4. 12 units |
34. To get to school from Bob's house you would have to travel in which direction?
- | | |
|----------|---------|
| 1. North | 3. East |
| 2. South | 4. West |
35. If you were at the Grocery Store and went south 4 units, where would you be?
- | | |
|----------------------|----------------|
| 1. Bike Shop | 3. Book Shop |
| 2. Old Haunted House | 4. Jim's House |
36. Suppose Jim left his house for a walk and went east 6 units, then south 4 units and then west 2 units. Where would he now be?
- | | |
|------------------|-----------------|
| 1. Record Shop | 3. Book Shop |
| 2. Grocery Store | 4. Garden Store |
37. Pete told you he left school and took the following path from school: A) East 10 units; B) South 4 units; C) West 6 units; D) South 4 units; E) East 6 units. Where is he now?
- | | |
|----------------|----------------|
| 1. Garden Shop | 3. Tim's House |
| 2. Bike Shop | 4. Book Shop |

* * * * *

Items 38-41 refer to the following pictures & explanations:



PICTURE A
BEFORE



PICTURE B
AFTER

Picture A shows a board with 3 nails before the start of an experiment. Picture B shows the same board after each nail was hit three times by the same person using the same hammer.

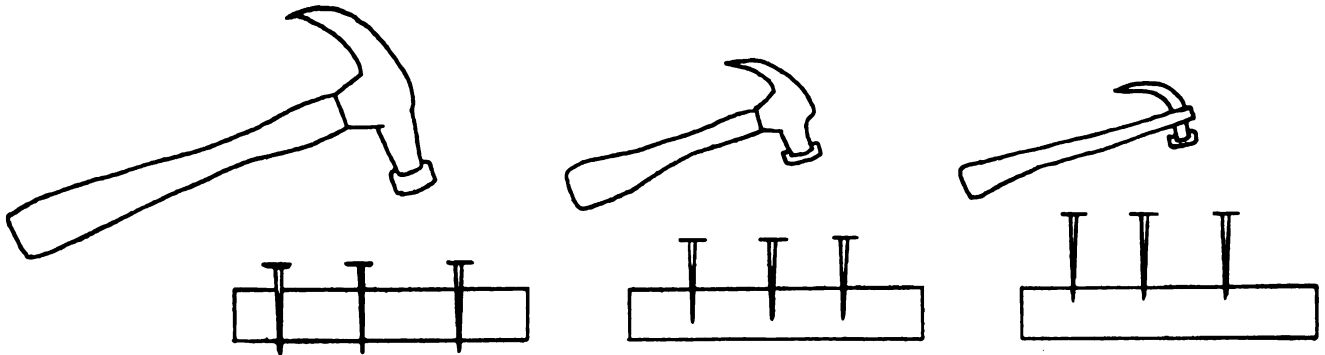
38. The energy source would be the

- | | |
|--------------------|--------------------|
| 1. hammer | 3. wood |
| 2. large nail only | 4. small nail only |

39. Which variable caused the results shown in Picture B?

1. The size of the hammer
2. The thickness of the nail
3. The thickness of the wood
4. The number of times each nail was hit

A student, using 3 different hammers to hit the nails 3 times each, had the following results shown in these pictures:

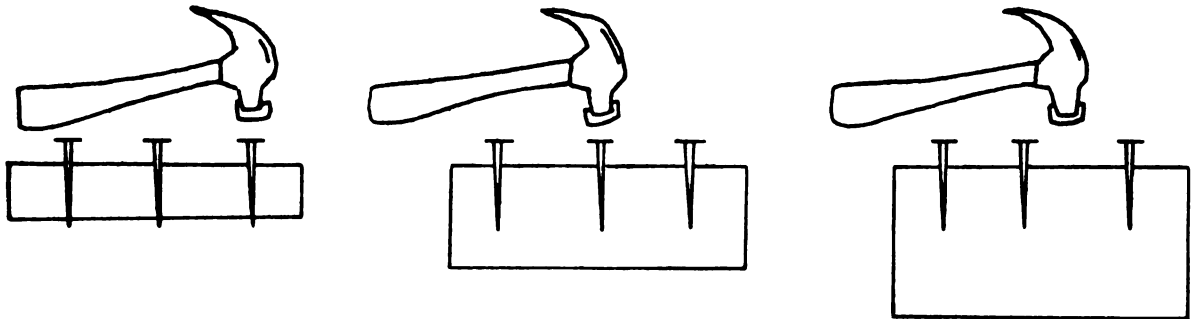


40. Which variable most likely caused the results shown in these pictures?

1. The hammers that were used
2. The size of the nails
3. The thickness of the wood
4. The number of times each nail was hit

* * *

Another experiment was tried with a student using the same hammer and hitting each nail 3 times. The results are shown below:

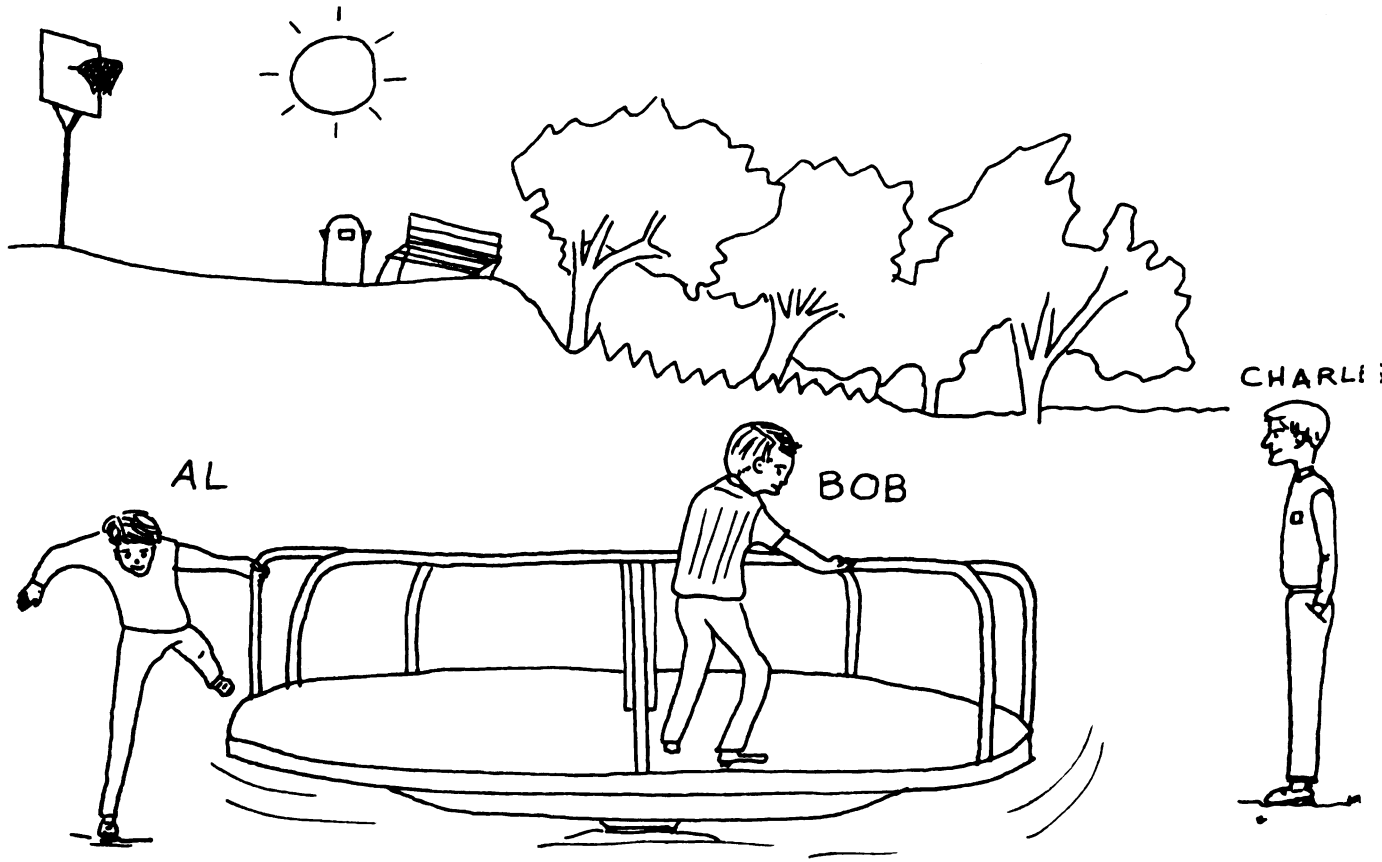


41. Which variable probably accounts for the results shown in these pictures?

1. The size of the hammer
2. The size of the nails
3. The thickness of the wood
4. The number of times each nail was hit

* * * * *

Items 42-44 refer to the following diagram and explanation:



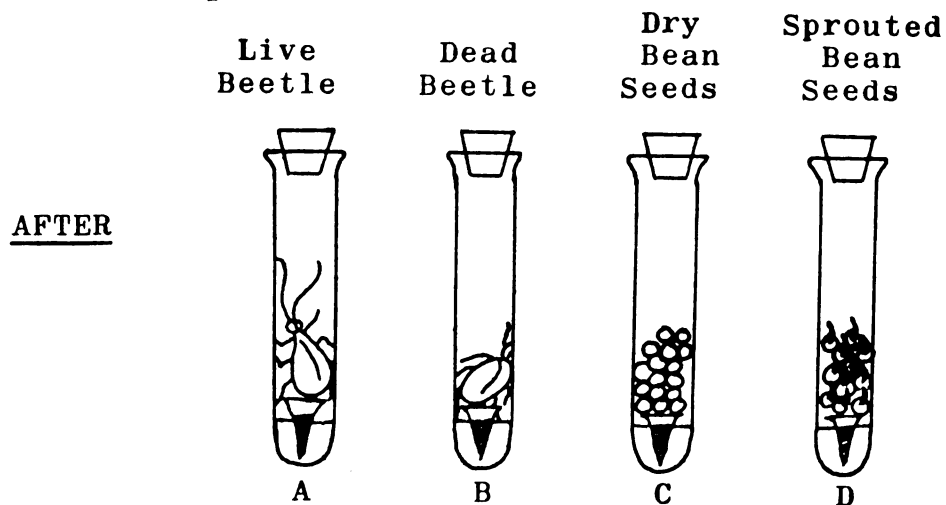
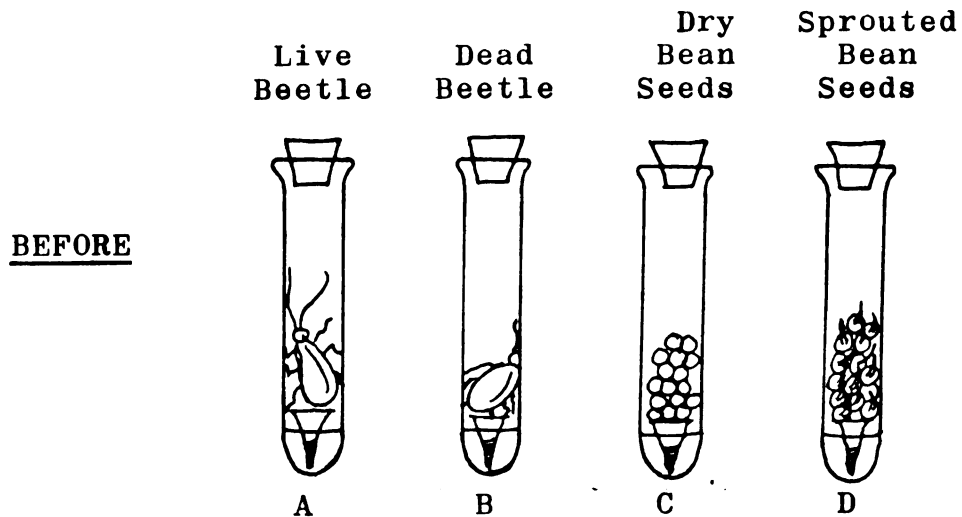
Study the diagram of the children playing with the playground merry-go-round. Locate the following children: Al, Bob, and Charles. Try to decide who could have made each of the following statements on the next page:

42. "I keep running but my position stays the same."
1. Al
 2. Bob
 3. Charles
 4. Al or Charles
43. "My distance to the park bench stays the same."
1. Al
 2. Bob
 3. Charles
 4. Al or Bob
44. "If you keep moving away from me, you are going to fall off the edge."
1. Al
 2. Bob
 3. Charles
 4. Al or Charles

* * * * *

Items 45-47 refer to the following experiment:

Containers were set up as shown in the diagram. Each container had a clean brass screw in the bottom, a few drops of a red dye (which turns yellow under certain conditions), and a cork stopper. Other things that were added are listed above the containers. After $\frac{1}{2}$ hour, the red dye in Containers A and D had turned yellow; the red dye in Containers B and C remained red.

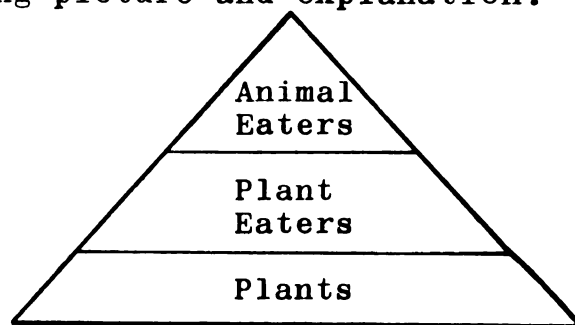


45. Containers A and B tell us that
1. All beetles cause the red dye to turn yellow
 2. Live beetles cause the red dye to turn yellow
 3. Beetles have no effect on the red dye
 4. Live beetles and dry bean seeds have no effect on the red dye
46. Containers A and D both resulted in a change of color because
1. They are on both ends of the experiment
 2. The containers were next to each other
 3. They both contain alive and growing organisms
 4. These organisms were trying to get out
47. If a container with a moving snail were part of the experiment, which container would it most look like after $\frac{1}{2}$ hour?
1. Container A
 2. Container B
 3. Container C
 4. Container D

* * * * *

For items 48-50 use the following picture and explanation:

This triangle shows the number of living things of each kind found in a field.



48. If the number of plant eaters were to decrease, what would happen to the number of plants?
1. They would increase
 2. They would remain the same
 3. They would decrease
49. If all light were kept from this field, what would happen?
1. All living things would soon die.
 2. Things would continue as usual
 3. Only the plants would die
50. If some new living things were added to the group and the number of plant eaters decreased, this could be because the new individuals
1. were plants
 2. were plant eaters
 3. were animal eaters

* * * * *

SCHOOL _____ PHONE _____

ADDRESS _____

PRINCIPAL _____

Number of 5th Grades: _____

<u>Number of Students</u>	<u>Teacher's Name</u>	<u>Room #</u>
-		
-		
-		
-		
-		
-		
-		

Thier Study: (circle one)

May 15 16 17 18 19

22 23 24 25 26

Science Programs:

Presently have:

Have had in past:

IQ

Which test -

When administered -

Edition -

Available -

APPENDIX B

TEST OF SCIENCE INQUIRY SKILLS, SEQUENTIAL
TEST OF EDUCATIONAL PROGRESS, AND SCIENCE
AND SCIENTISTS ATTITUDE INVENTORY RAW
SCORE DISTRIBUTIONS FOR SCIS AND
CONTROL POPULATIONS

Table B1.--TSIS raw score distributions--SCIS population.

Raw Score	Frequency	Cumulative Frequency	Standard Score
47	3	3	70.2
46	3	6	69.2
45	1	7	68.2
44	3	10	67.2
43	11	21	66.2
42	5	26	65.2
41	13	39	64.2
40	9	48	63.2
39	5	53	62.1
38	11	64	61.1
37	13	77	60.1
36	11	88	59.1
35	9	97	58.1
34	8	105	57.1
33	9	114	56.1
32	13	127	55.1
31	8	135	54.1
30	9	144	53.1
29	9	153	52.1
28	11	164	51.0
27	11	175	50.0
26	8	183	49.0
25	9	192	48.0
24	5	197	46.0
23	7	204	45.0
22	11	215	45.0
21	13	228	44.0
20	5	233	43.0
19	12	245	42.0
18	12	257	41.0
17	8	265	40.0
16	12	277	38.9
15	8	285	37.9
14	3	288	36.9
13	1	289	35.9
12	4	293	34.9
11	5	298	33.9
10	7	305	32.9
9	2	307	31.9
8	3	310	30.9

Mean 26.91

Standard Deviation 9.91

Variance 98.27

Standard score has mean of 50 and standard deviation of 10

Table B2.--TSIS summary data--SCIS population.

DISTRIBUTION OF ITEM DIFFICULTY INDICES

	<u>Number of Items</u>	<u>Percentage</u>
91-100	0	0
81- 90	0	0
71- 80	1	2
61- 70	2	4
51- 60	11	22
41- 50	18	36
31- 40	11	22
21- 30	5	10
11- 20	2	4
0- 10	0	0

DISTRIBUTION OF DISCRIMINATION INDICES

	<u>Number of Items</u>	<u>Percentage</u>
91-100	0	0
81- 90	1	2
71- 80	2	4
61- 70	13	26
51- 60	16	32
41- 50	5	10
31- 40	5	10
21- 30	6	12
11- 20	1	2
0- 10	1	2

MEAN ITEM DIFFICULTY	46
----------------------	----

MEAN ITEM DISCRIMINATION	49
--------------------------	----

KUDER-RICHARDSON RELIABILITY #20	.9033
----------------------------------	-------

STANDARD ERROR OF MEASUREMENT	3.0810
-------------------------------	--------

Table B3.--TSIS raw score distributions--Control population.

Raw Score	Frequency	Cumulative Frequency	Standard Score
47	1	1	68.8
46	3	4	67.7
45	3	7	66.5
44	4	11	65.4
43	3	14	64.2
42	7	21	63.1
41	6	27	61.9
40	9	36	60.8
39	5	41	59.6
38	3	44	58.5
37	5	49	57.3
36	10	59	56.2
35	10	69	55.0
34	4	73	53.9
33	10	83	52.7
32	14	97	51.5
31	10	107	50.4
30	5	112	49.2
29	6	118	48.1
28	5	123	46.9
27	9	132	45.8
26	7	139	44.6
25	5	144	43.5
24	6	150	42.3
23	3	153	41.2
22	10	163	40.0
21	3	166	38.9
20	1	167	37.7
19	2	169	36.6
18	5	174	35.4
17	2	176	34.3
16	2	178	33.1
15	4	182	32.0
14	2	184	30.8
13	3	187	29.7
12	1	188	28.5
11	2	190	27.4
10	1	191	26.2

Mean 30.61

Standard Deviation 8.69

Variance 75.60

Standard score has mean of 50 and standard deviation of 10

Table B4.--TSIS summary data--Control population.

DISTRIBUTION OF ITEM DIFFICULTY INDICES

	<u>Number of Items</u>	<u>Percentage</u>
91-100	0	0
81- 90	0	0
71- 80	2	4
61- 70	0	0
51- 60	7	14
41- 50	14	28
31- 40	13	26
21- 30	11	22
11- 20	3	6
0- 10	0	0

DISTRIBUTION OF DISCRIMINATION INDICES

	<u>Number of Items</u>	<u>Percentage</u>
91-100	0	0
81- 90	0	0
71- 80	1	2
61- 70	6	12
51- 60	13	26
41- 50	11	22
31- 40	9	18
21- 30	5	10
11- 20	4	8
0- 10	1	2

MEAN ITEM DIFFICULTY	39
MEAN ITEM DISCRIMINATION	43
KUDER-RICHARDSON RELIABILITY #20	.8726
STANDARD ERROR OF MEASUREMENT	3.1017

Table B5.--STEP raw score distributions--SCIS population.

Raw Score	Frequency	Cumulative Frequency	Standard Score
45	2	2	71.2
44	2	4	69.9
43	3	7	68.6
42	3	10	67.4
41	6	16	66.1
40	6	22	64.9
39	12	34	63.6
38	6	40	62.3
37	17	57	61.1
36	10	67	59.8
35	12	79	58.5
34	16	95	57.3
33	12	107	56.0
32	15	122	54.8
31	12	134	53.5
30	9	143	52.2
29	15	158	51.0
28	16	174	49.7
27	17	191	48.4
26	15	206	47.2
25	9	215	45.9
24	16	231	44.7
23	16	247	43.4
22	10	257	42.1
21	11	268	40.9
20	6	274	39.6
19	10	284	38.3
18	9	293	37.1
17	1	294	35.8
16	6	300	35.4
15	8	308	33.3
14	5	313	32.0
13	4	317	30.8
11	5	322	28.2
10	2	324	27.0

Mean 28.23

Standard Deviation 7.92

Variance 62.73

Standard score has a mean of 50 and standard deviation of 10

Table B6.--STEP raw score distributions--Control population.

Raw Score	Frequency	Cumulative Frequency	Standard Score
49	1	1	73.4
46	4	5	70.1
45	1	6	69.0
44	1	7	67.9
43	1	8	66.8
42	3	11	65.7
41	5	16	64.5
40	5	21	63.4
39	4	25	62.3
38	5	30	61.2
37	6	36	60.7
36	5	41	59.0
35	11	52	57.9
34	4	56	56.8
33	8	64	55.7
32	6	70	54.5
31	6	76	53.4
30	6	82	52.3
29	6	88	51.2
28	7	95	50.1
27	5	100	49.0
26	5	105	47.9
25	7	112	46.8
24	9	121	45.7
23	8	129	44.5
22	2	131	43.4
21	10	141	42.3
20	9	150	41.2
19	8	158	40.1
18	4	162	39.0
17	6	168	37.9
16	4	172	36.8
15	3	175	35.7
14	1	176	34.5
13	2	178	33.4
12	4	182	32.3
11	2	184	31.2
8	2	186	27.9

Mean 27.91

Standard Deviation 9.00

Variance 81.01

Standard score has mean of 50 and standard deviation of 10

Table B7.--SASAI raw score distributions--SCIS population.

Raw Score	Frequency	Cumulative Frequency	Standard Score
48	4	4	66.6
47	3	7	65.2
46	3	10	63.8
45	16	26	62.4
44	13	39	61.0
43	15	54	59.6
42	24	78	58.2
41	31	109	56.8
40	17	126	55.4
39	24	150	53.9
38	15	165	52.5
37	15	180	51.1
36	16	196	49.7
35	15	211	48.3
34	13	224	46.9
33	11	235	45.5
32	18	253	44.1
31	9	262	42.7
30	11	273	41.3
29	9	282	39.9
28	4	286	38.5
27	8	294	37.1
26	4	298	35.7
25	2	300	34.3
24	3	303	32.8
23	1	304	31.4
22	6	310	30.0
21	4	314	28.6
20	2	316	27.2
18	5	321	24.4
17	1	322	23.0
16	1	323	21.6
10	1	324	13.2
8	1	325	10.3

Mean 36.16

Standard Deviation 7.11

Variance 50.58

Standard score has mean of 50 and standard deviation of 10

Table B8.--SASAI raw score distributions--Control population.

Raw Score	Frequency	Cumulative Frequency	Standard Score
49	1	1	68.1
48	2	3	66.7
47	2	5	65.3
46	4	9	63.9
45	12	21	62.5
44	3	24	61.1
43	9	33	59.7
42	11	44	58.3
41	18	62	56.9
40	7	69	55.4
39	13	82	54.0
38	15	97	52.6
37	8	105	51.2
36	11	116	49.8
35	11	127	48.4
34	6	133	47.0
33	6	139	45.6
32	6	145	44.2
31	5	150	42.8
30	8	158	41.4
29	5	163	40.0
28	1	164	38.6
27	6	170	37.2
26	3	173	35.8
25	7	180	34.4
24	1	181	33.0
23	2	183	31.6
22	2	185	30.2
19	4	189	26.0
18	1	190	24.6
17	1	191	23.2
11	1	192	14.8

Mean 36.08

Standard Deviation 7.13

Variance 50.92

Standard score has mean of 50 and standard deviation of 10

APPENDIX C

PRELIMINARY FORMS A AND B,
TEST OF SCIENCE INQUIRY SKILLS

A TEST OF SCIENCE INQUIRY SKILLS

Preliminary Edition by Joseph W. Riley

DIRECTIONS

1. Remove the answer sheet from the test booklet. Close the booklet after removing the answer sheet.
2. Print your name, last name first, in the space provided in the upper left corner of the answer sheet.
3. In the space next to your name, write your teacher's name.
4. In the spaces under your name, write your birth date, for example, January 4, 1961.
5. In the boxes in the upper right corner print in as much of your name as possible and fill in the correct spaces under your name.
6. Find the box labeled "sex" at the bottom of the page. If you are a boy, fill in the space labeled "M". If you are a girl, fill in the space labeled "F".
7. Now turn the answer sheet so that your name is on the left side of the page.
8. Find Number 1 on the answer sheet, then find Number 8. Notice that the numbers run ACROSS the page, from left to right.
9. Now open your test booklet. Each question in this test is followed by 3 or 4 choices for an answer. Read each question and then decide which is the one best answer.
10. On the answer sheet find the row which has the same number as the question you are answering. In that row blacken the space that has the number which is the same as the one in front of the answer you choose.
11. Make your marks heavy and black, press down hard on your pencil, and go over your marks. If you make a mistake, erase completely the answer you wish to change. Use only the special pencil that has been provided.
12. Look at this sample question:

85. A cow is a:

1. machine 2. plant 3. animal 4. rock

For this question, the answer sheet should look like this:

85. ☐ 1 ☐ 2 ☒ 3 ☐ 4

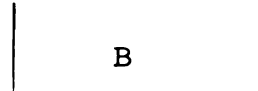
For each item in this examination select the ONE best answer.
Do not mark more than ONE answer space for each item.

* * * * *

Items 1-4 refer to the following diagrams: (Remember that an organism can be any individual plant or animal.)



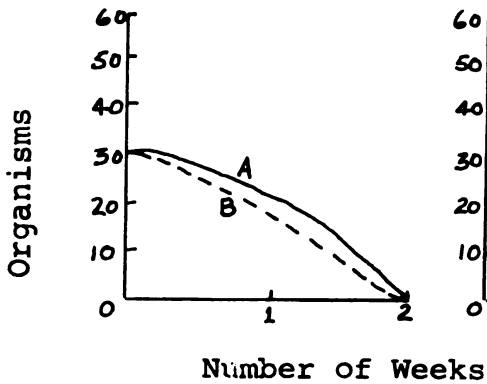
Before



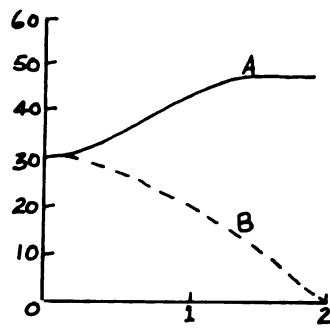
2 Weeks Later

1. If you placed equal numbers of two kinds of organisms (A and B) in an aquarium and 2 weeks later only type B was left, the best explanation would be:
 1. The aquarium was suitable for A and B
 2. The aquarium was unsuitable for A and B
 3. The aquarium was suitable for B but not A
 4. The aquarium was suitable for A but not B
2. Since there were no other organisms except A and B at the beginning, we might expect that
 1. Organism A ate Organism B
 2. Organism B ate Organism A
 3. A and B avoided each other
 4. All the B individuals got away

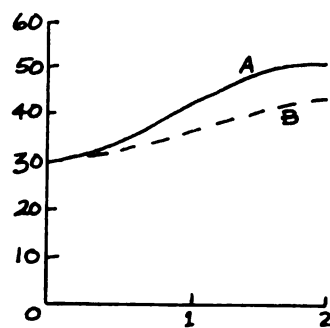
GRAPH I



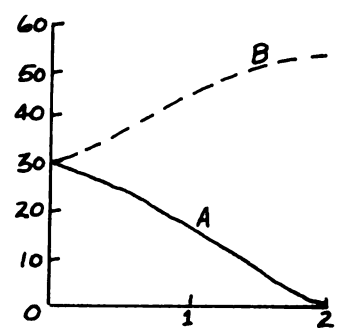
GRAPH II



GRAPH III



GRAPH IV



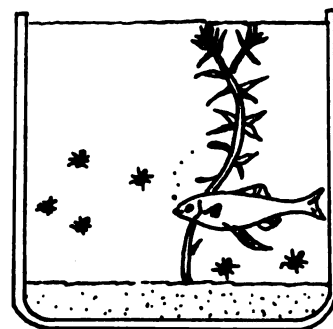
3. Which graph would best show what happened in the aquarium?

(The solid line represents Organism A and the dotted line represents Organism B.)

1. Graph I 2. Graph II 3. Graph III 4. Graph IV

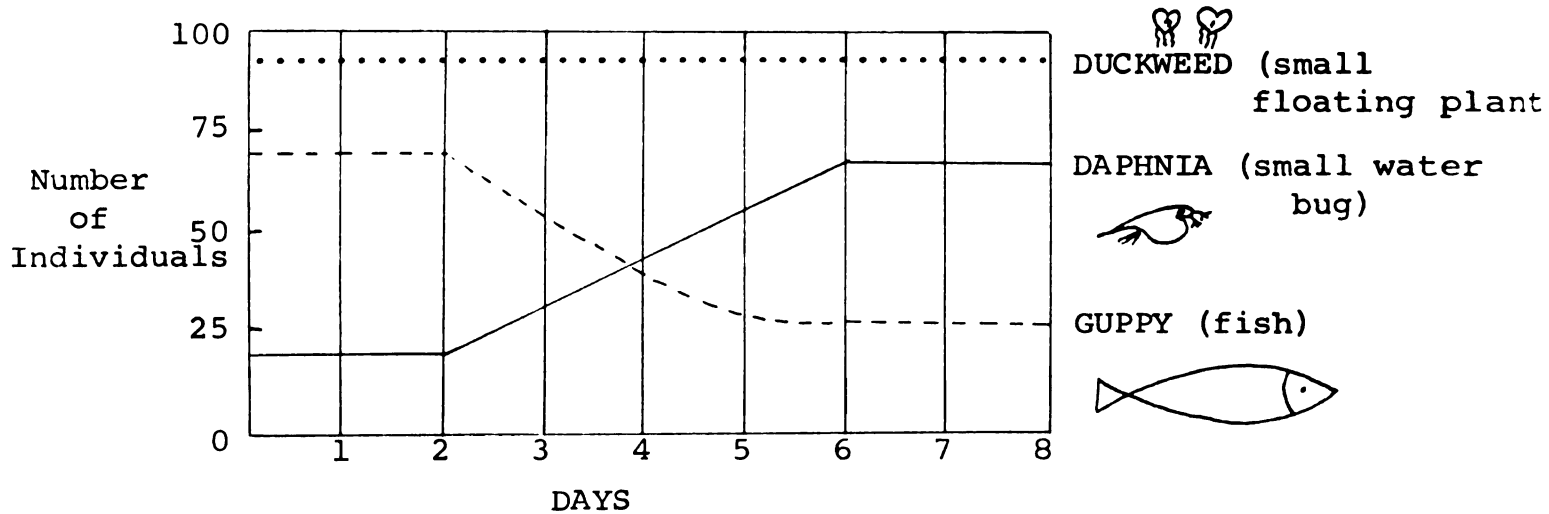
4. If Organism A were a small bug and Organism B were a large goldfish and the aquarium was placed in the dark, the decrease in Organism A could have been caused by

1. A eating B
2. B eating A
3. A died because they had no light
4. All 3 of these answers are correct



* * * * *

Items 5 - 11 refer to what happened in an aquarium as shown by the graph. NOTE: Some guppies were taken out each day after the second day.

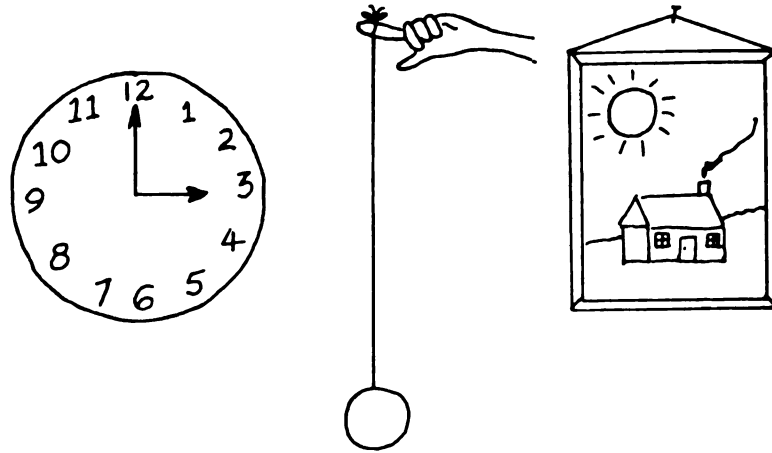


5. During this experiment the number of Duckweed plants
 1. increases
 2. decreases
 3. stays the same
6. During this experiment the number of Daphnia
 1. increases
 2. decreases
 3. stays the same
7. During this experiment the number of Guppies
 1. increases
 2. decreases
 3. stays the same
8. What happened to all the populations during this experiment?
 1. Duckweed decreased, Guppies increased
 2. Guppies decreased, Daphnia increased
 3. Daphnia decreased, Guppies increased

9. The change in the Guppy population after the second day is due to:
 1. a disease in the aquarium
 2. the Guppies were taken out of the tank
 3. the Guppies stopped having babies
 4. the food supply ran out
10. The change in the Daphnia population may be due to:
 1. a decrease in the number of Duckweed plants
 2. an increase in the number of Duckweed plants
 3. a decrease in the number of Guppies
 4. an increase in the number of Guppies
11. The populations causing the change are:
 1. Daphnia and Duckweed
 2. Guppies and Duckweed
 3. Guppies and Daphnia

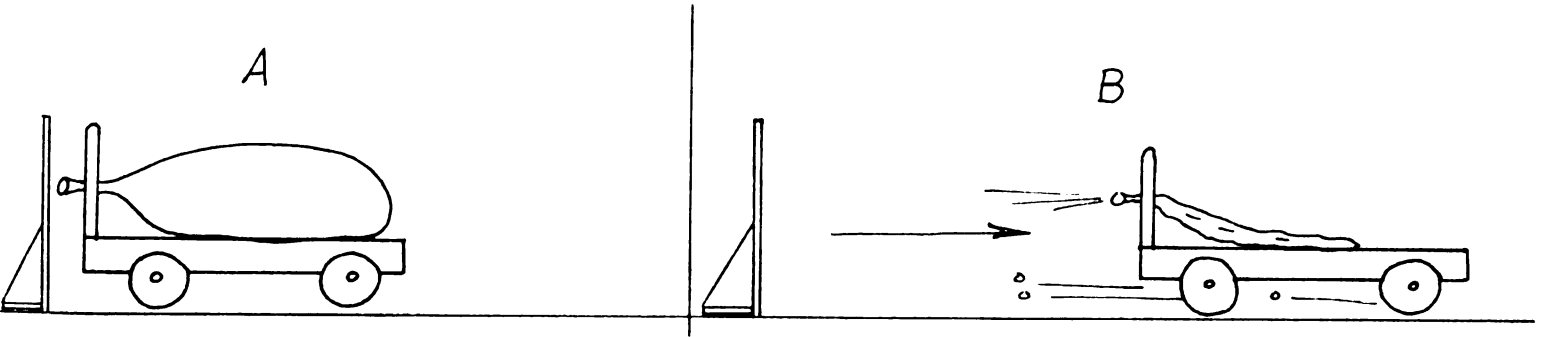
* * * * *

Items 12 and 13 refer to the following diagram:



12. Look at the swing system pictured above. What are the parts of this swing system?
1. ball, string, finger
 2. picture, string, ball
 3. picture, ball, string, finger
 4. clock, picture, finger
13. Suppose you are investigating how long it takes the ball to swing back and forth. What variable in the system would make a difference in the time?
1. The strength of the person who is holding the string.
 2. The length of the string.
 3. The surface of the ball (whether it is rough or smooth)
 4. The time on the clock

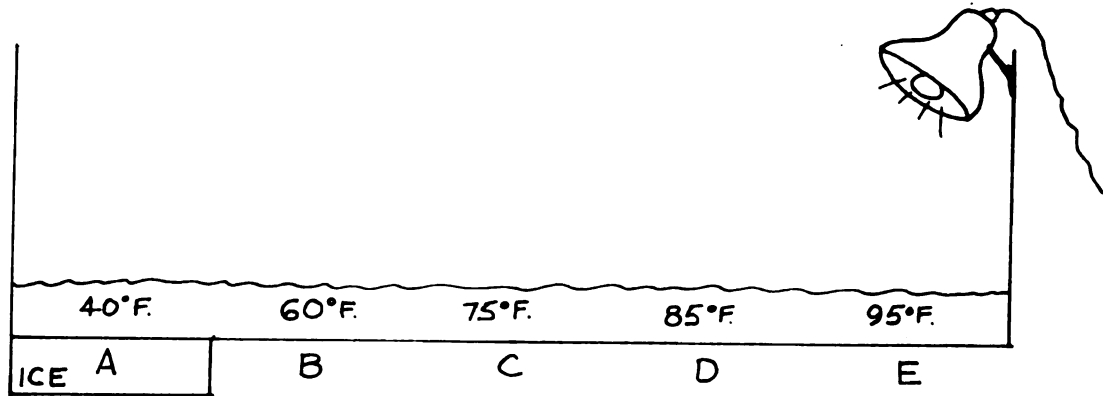
Items 14-16 refer to the following diagrams:



14. What is the immediate source of energy in this system?
1. the wheels
 2. the air in the balloon
 3. the wall
 4. the cart
15. Is the system in Picture B the same system as in Picture A?
1. Yes
 2. No
 3. Cannot be determined
16. The evidence that interaction is taking place is:
1. the balloon is full of air in Picture A
 2. the cart moves
 3. the balloon remains on the cart

* * * * *

Items 17-24 refer to the following diagrams and explanations:



The above container is set on a table in the classroom. The entire container is exposed to daylight and the E end has a 200 watt light bulb mounted over it. This light will be left on all of the time. The bottom of the container was covered with 2 inches of potting soil. Water was sprinkled over the entire surface of the container two times a week. Container A was full of ice all the time. After 3 days the temperatures were taken at each section of the container and recorded as follows:

A--40°F. B--60°F. C--75°F. D--85°F. E--95°F.

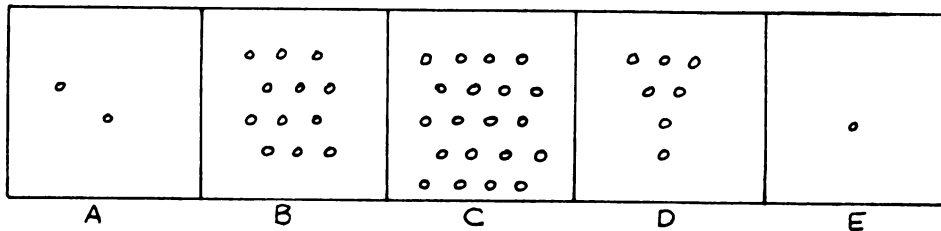
17. Section E was warmer because

1. it was closer to the window
2. the light was at that end
3. the soil was deeper at that end

18. You will recall that all portions of the container received equal amounts of water 2 times a week. Still, Section E was much drier than Section A. This can be explained because

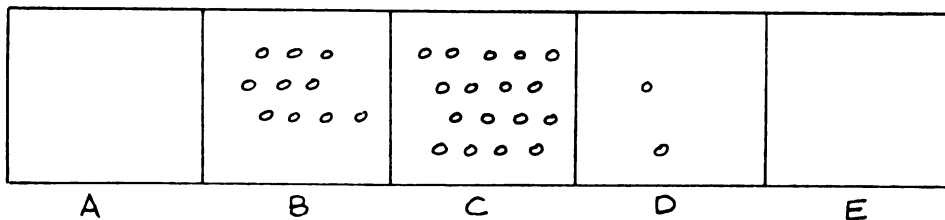
1. Section E was closer to the window
2. the light was at that end
3. the soil was deeper at that end

19. 100 clover seeds were planted, 20 in each section. One week later the number of clover plants that could be seen above the ground were counted and recorded on this chart.



From this recorded data, where did most of the clover plants grow?

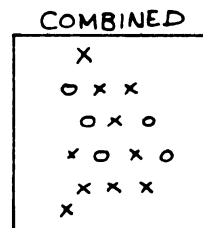
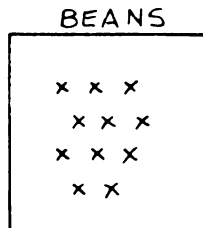
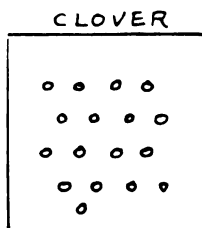
1. Section A 2. Section B 3. Section C 4. Section D
20. 3 weeks after planting, the number of clover plants that were still growing were counted and recorded on this chart.



From this recorded data, you can conclude that

1. Once clover starts to grow, it always continues growing
 2. Seeds don't always grow in environments where they sprout
 3. Clover seeds grow best in the darkest and coolest place
21. What factor may have been responsible for the growth pattern of the clover?
1. light
 2. temperature
 3. moisture
 4. all of these

The following chart shows what happened after 3 weeks when 20 clover seeds were planted alone in Section C, 20 bean seeds planted alone in Section C, and 20 each of bean and clover grown together in Section C.



NOTE: the soil felt drier when both seeds were planted together

22. When clover and beans are grown together the number of clover plants

1. increases
2. decreases
3. stays the same

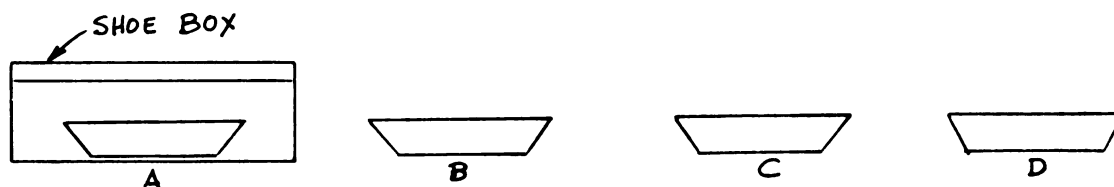
23. A reason for this could be

1. the bean plants shaded the clover and used up much of the water in the soil
2. the bean plants crowded out the clover
3. the bean plants shaded the clover
4. the clover plants shaded the bean plants

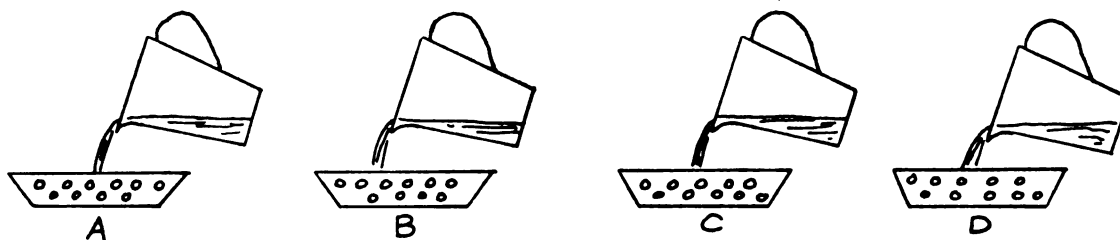
Suppose we wanted to find out what effect moisture had on the number of clover plants that could grow for 3 weeks. The following materials were used: 4 aluminum pie pans; 40 seeds; water pitcher with measuring cup.

24. Which of the following things would you do to find out if the amount of water had any effect on the number of clover seeds that could grow for 3 weeks?

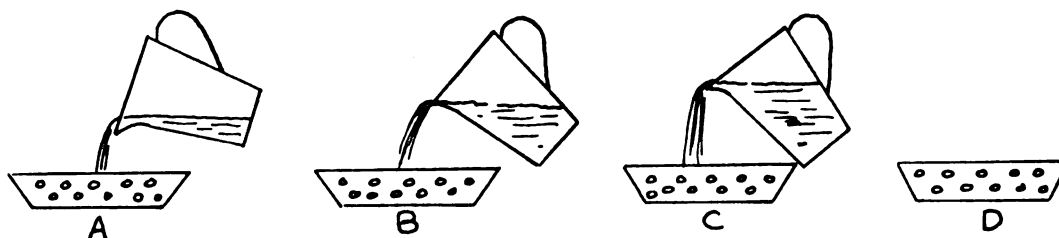
1. Place 1 container in a shoe box & leave the other 3 in light.



2. Place 10 seeds in each container and add 1 cup of water to each container. Leave each container out in the room.

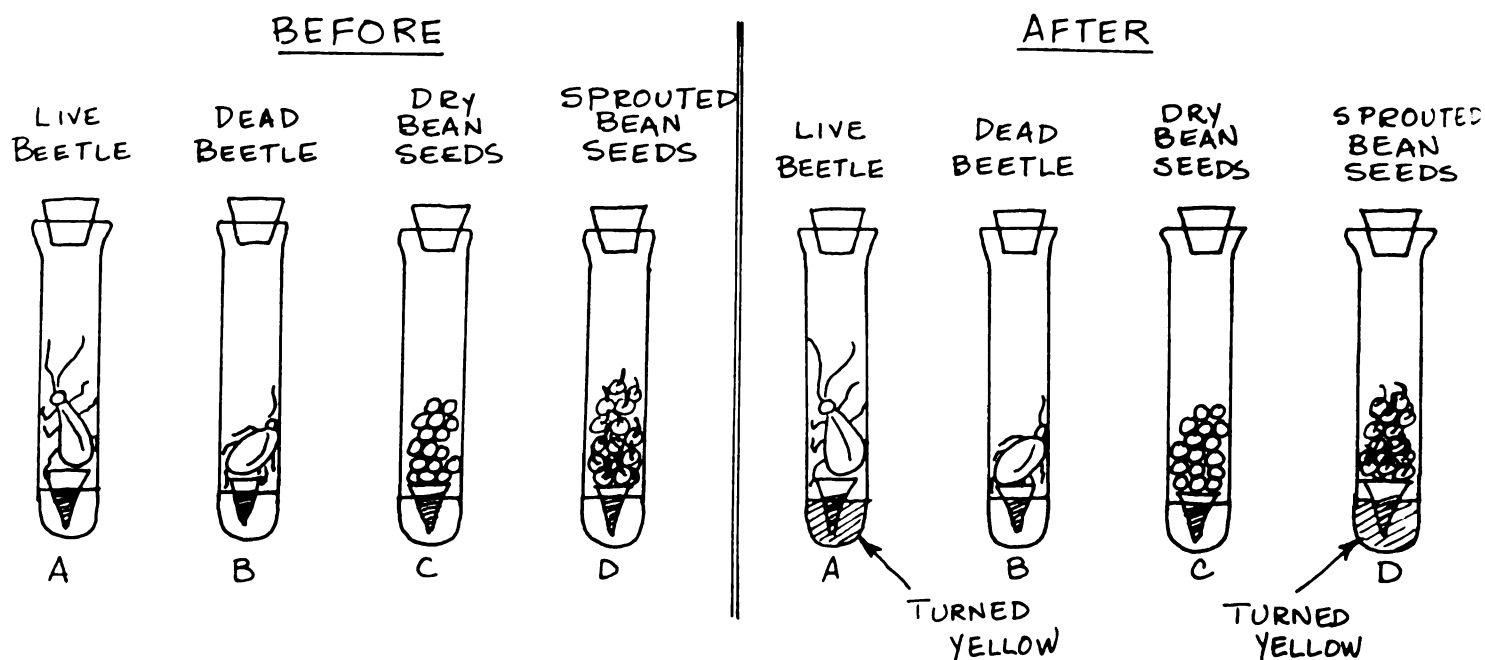


3. Place 10 seeds in each container. Add 1 cup of water to Pan A, 2 cups of water to Pan B, 3 cups of water to Pan C, and no water to Pan D. Leave each container out in the room.



Items 25-30 refer to the following experiment:

Containers were set up as shown in the diagram. Each container had a clean brass screw in the bottom, a few drops of a red dye (which turns yellow under certain conditions), and a cork. Other contents are listed above the containers. After 1/2 hour, the red dye in Containers A and D had turned yellow; the red dye in Containers B and C remained red.



25. Containers A and B tell us that

1. All beetles cause the red dye to turn yellow
2. Live beetles cause the red dye to turn yellow
3. Beetles have no effect on the red dye
4. Live beetles and dry bean seeds have no effect on the red dye

26. There was no change in color in Containers B and C because

1. The red dye in containers B and C was no good
2. Neither of these things was growing
3. Plants will not cause the red dye to change
4. The brass screw did not work

27. Containers A and D both resulted in a change of color because
1. They are on both ends of the experiment
 2. They both contain alive and growing organisms
 3. The containers were next to each other
 4. These organisms were trying to get out
28. If a container with an empty snail shell were part of the experiment, which container would it most resemble after $\frac{1}{2}$ hour?
1. Container A
 2. Container B
 3. Container C
 4. Container D
29. If a container with a moving snail were part of the experiment, which container would it most resemble after $\frac{1}{2}$ hour?
1. Container A
 2. Container B
 3. Container C
 4. Container D
30. What could be done to this experiment to make it better?
1. Take away the container with the live beetle
 2. Add a fifth container that only has a clean brass screw, a few drops of the red dye, and a cork
 3. Add a fifth container that would be the same as Container A
 4. All of these would improve this experiment.

* * * *

Items 31-35 refer to the following experiment:

Three unknown powders labeled A, B, and C are tested.

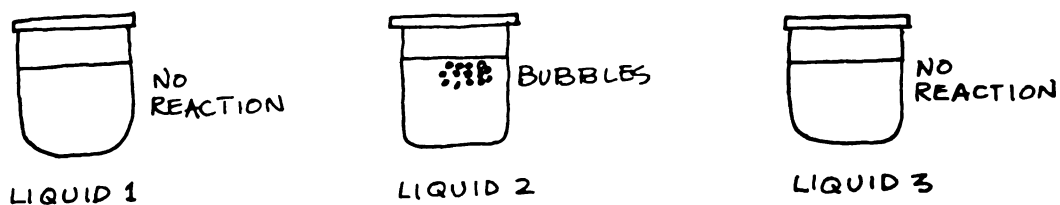
STEP I: A pinch of Powder A is put into each of three jars containing different colorless liquids and labeled Liquid 1, 2, and 3. (See diagram)

STEP II: Next, a pinch of Powder B is put into a second set of jars labeled Liquid 1, 2, and 3.

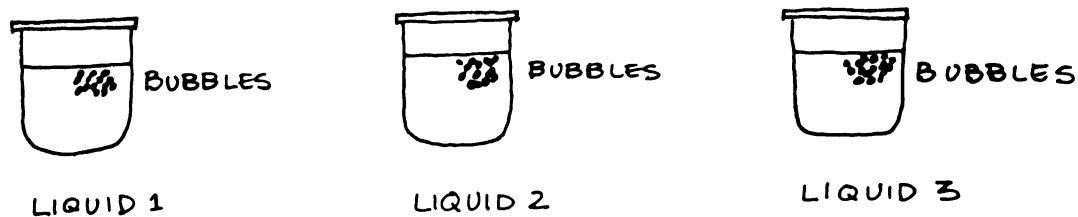
STEP III: Finally, small amounts of Powder C are put into a third set of jars labeled Liquid 1, 2, and 3. The results are shown in the diagram below.

DIAGRAM I

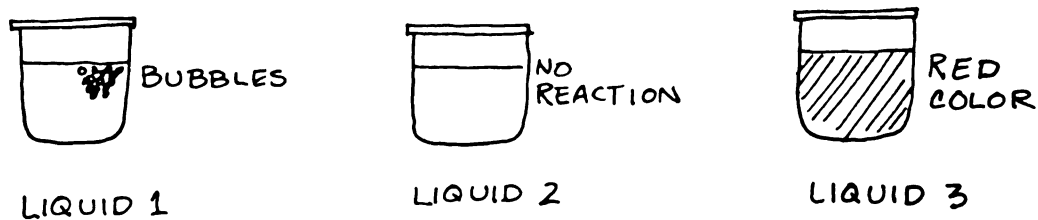
POWDER A:



POWDER B:



POWDER C:



31. Suppose Powder A were added to Mystery Liquid Q producing no reaction. From this result, Mystery Liquid Q could be
1. Liquid 1 or 3
 2. Liquid 1, 2, or 3
 3. Liquid 3 only
 4. Liquid 2 only
32. Suppose Powder C was added to Mystery Liquid Q producing a red color. Mystery Liquid Q could be
1. Liquid 1 or 3
 2. Liquid 1, 2, or 3
 3. Liquid 3 only
 4. Liquid 2 only
33. Suppose Powder B and Mystery Liquid R produced bubbles. We could conclude that Mystery Liquid R could be
1. Liquid 1, 2, or 3
 2. Liquid 1 or 2 only
 3. Liquid 2 only
 4. Liquid 3 only
34. Suppose Powder C were added to Mystery Liquid R producing no reaction. We could conclude that Mystery Liquid R could be
1. Liquid 1, 2, or 3
 2. Liquid 1 or 2
 3. Liquid 2 only
 4. Liquid 3 only
35. Suppose a mystery powder was added to each of the liquids and the following results were produced:
- Mystery Powder X + Liquid 1 produced bubbles
- Mystery Powder X + Liquid 2 produced no reaction
- Mystery Powder X + Liquid 3 produced a red color
- From these results, the mystery powder X is probably:
1. Powder A
 2. Powder B
 3. Powder C
 4. None of these

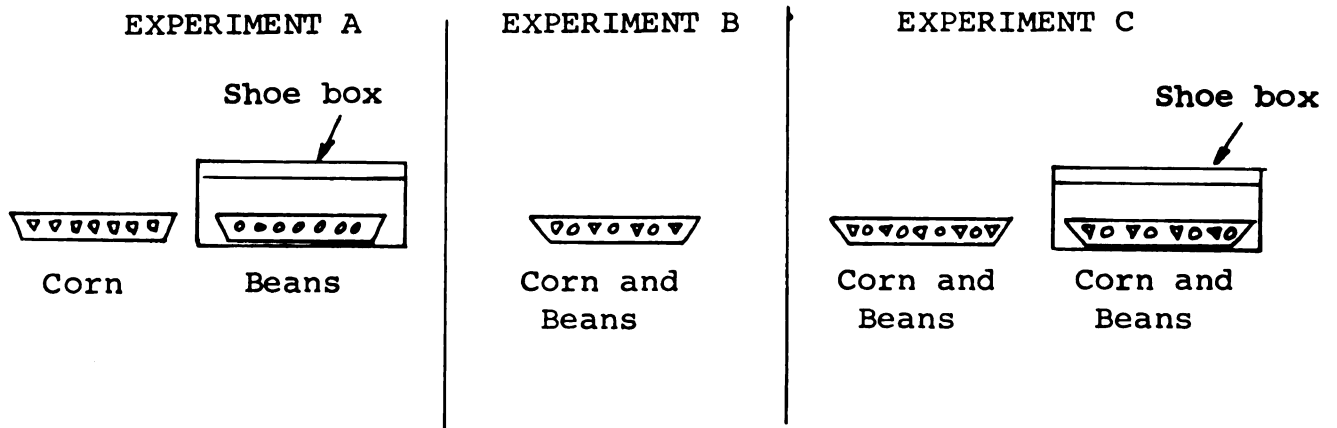
Items 36 and 37 refer to the following experiment and diagrams:

Suppose you had the problem of finding out what light does to corn and bean plants. Three possible answers are:

- 1) Light causes plants to grow tall, thick and healthy
- 2) Light causes plants to grow tall and weak
- 3) Light causes no change in the height of plants

Look at the three experiments drawn below:

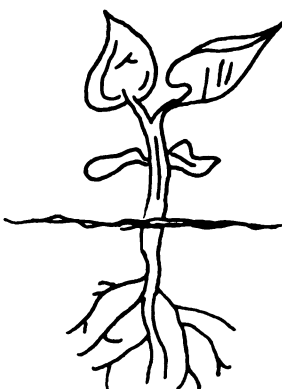
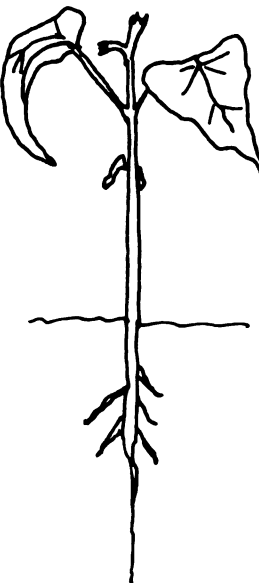
▽ = Corn seeds ○ = Bean seeds



36. Which set-up would you use to help solve the problem?

1. Experiment A
2. Experiment B
3. Experiment C

Assume that a similar experiment was tried, with the following results:

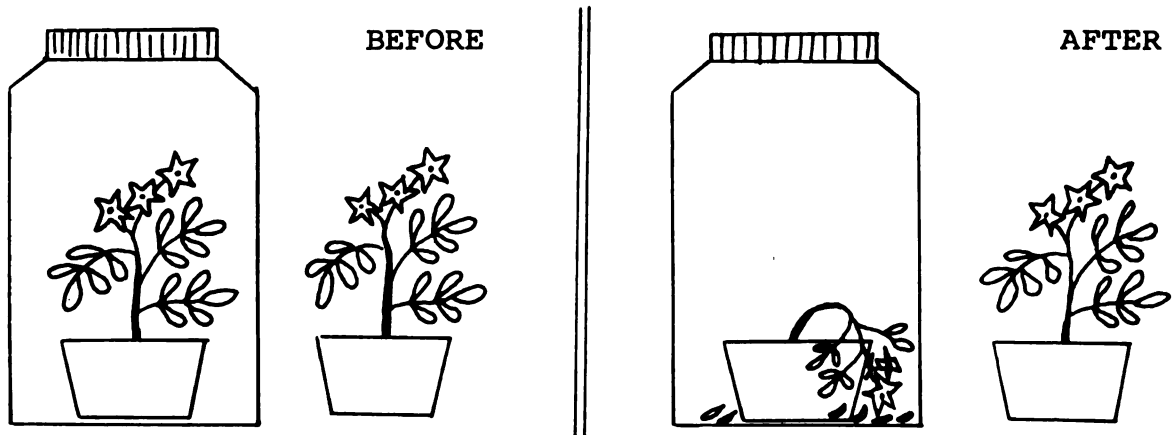
	LIGHT	DARK
CORN	Leaves large Leaves spread out Short plants 	Leaves small Leaves not spread out Tall plants 
BEAN	Stems thick Stems short 	Stems skinny Stems tall 

37. Based on these results, we can conclude that

1. Light has no effect on the height of plants
2. Plants grown in the dark are taller than plants grown in the light
3. Plants grown in the light are taller than plants grown in the dark

Items 38-41 refer to the following explanations and diagrams:

Suppose you place a plant in a sealed jar and a smaller plant beside it in the open air. Study the diagram.

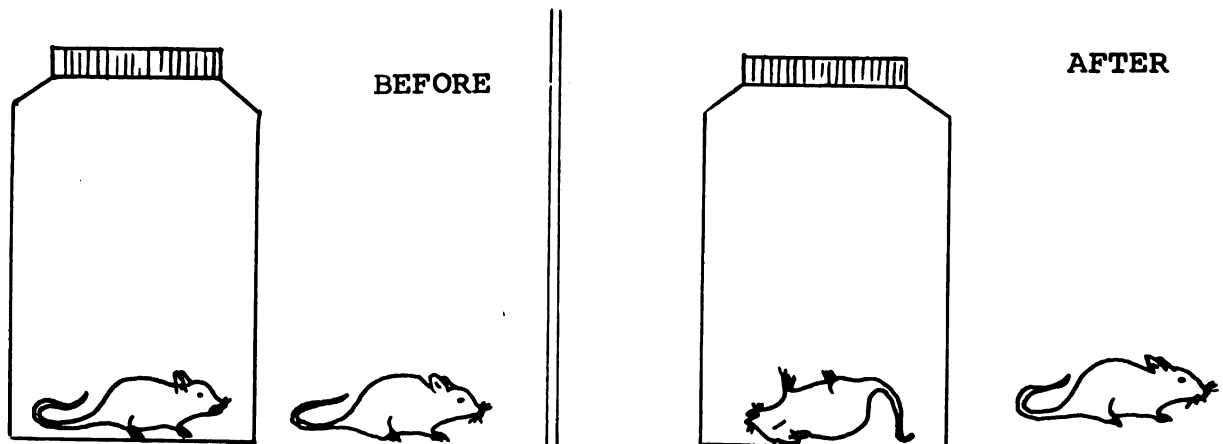


After some time the plant in the jar dies. The one planted in the open air continues to grow.

38. From the data presented thus far you can be absolutely sure that:

1. the plant in the jar uses up something in the air that it needs to live
2. the plant in the jar produces something that pollutes the air and kills it
3. Both explanations could be correct

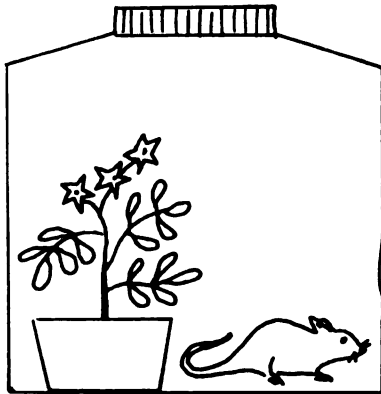
If two mice were used with one mouse placed alone in the sealed jar and one out in the open, the mouse in the sealed jar soon dies.



39. From this data you can conclude that:

1. the mouse in the jar uses up something in the air that it needs to live
2. The mouse in the jar produces something that pollutes the air and kills it
3. Both explanations could be correct

40. Now suppose that the plant and mouse are placed in a sealed jar together.



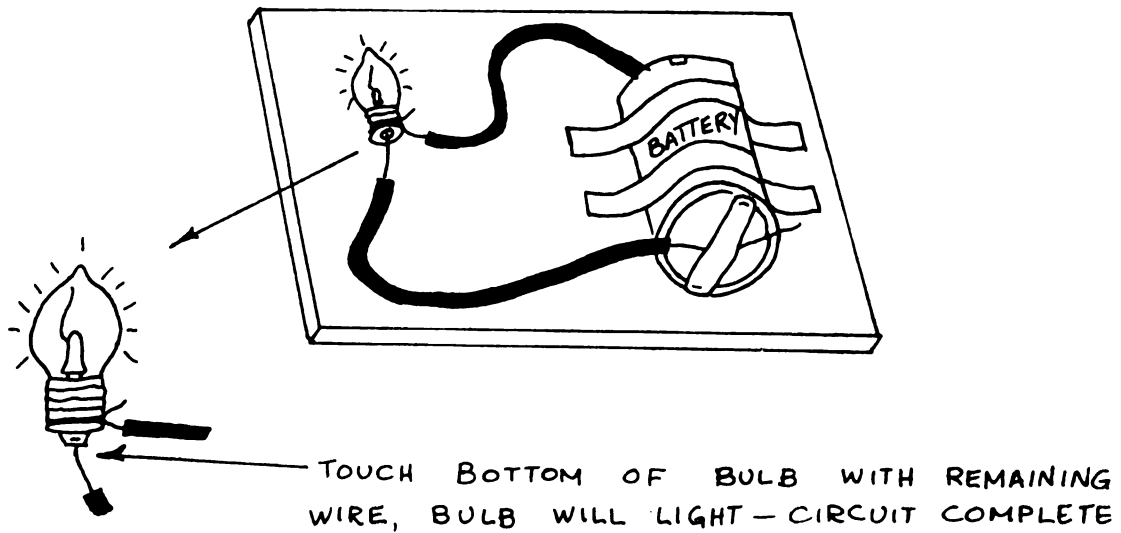
Based ONLY on the evidence presented thus far you would predict that:

1. the mouse and plant would die sooner than when they were grown in the jar separately
2. the plant would live but the mouse would die
3. the plant and mouse would live longer than if they were grown separately

41. When this experiment was tried, if it was found that the plant and mouse both lived longer than when they were in the jar separately, how could this best be explained?

1. They both use the same material out of the air
2. They both produce the same polluting substance and release it into the air.
3. They each produce and use different substances. Therefore, they help each other out.
4. All 3 explanations could be correct.

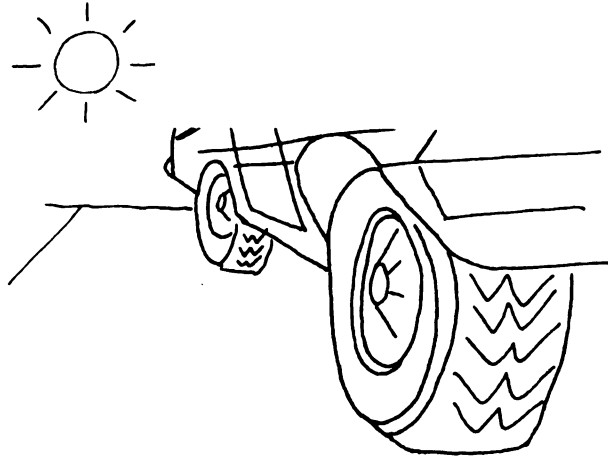
Items 42 and 43 refer to the following diagram:



42. What is the source of energy in this system?
1. light bulb
 2. wires
 3. battery
 4. tape
43. What is the evidence that interaction is taking place?
1. the wires are curved
 2. the light bulb lights
 3. the battery is taped down

* * * * *

Items 44 and 45 refer to the following diagram and explanation:



After a tire has been travelling over the road for a few minutes it was noted that the temperature of the tire is higher than that of the road. It has also been noted that if a car motor is running but not moving for the same period of time, that the tire temperature does not change.

44. Where did the heat come from?

- | | |
|---------------|--------------------------------------|
| 1. the engine | 3. the tire pushing against the road |
| 2. the sun | 4. the gasoline |

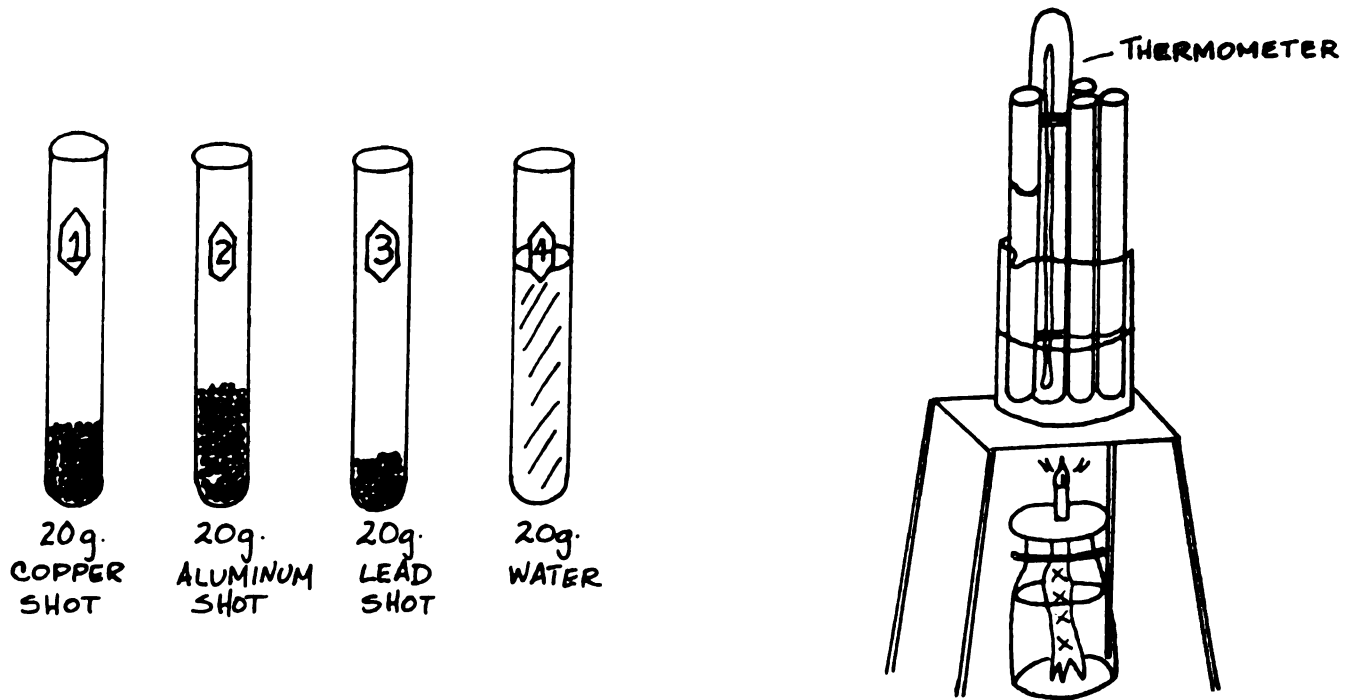
45. What was the energy source?

- | | |
|---------------|----------------------------------|
| 1. the engine | 3. the air in the tire |
| 2. the sun | 4. the steering wheel of the car |

* * * * *

Items 46-52 refer to the following experiment:

The four test tubes were placed in a beaker of boiling water with a thermometer for 5 minutes.



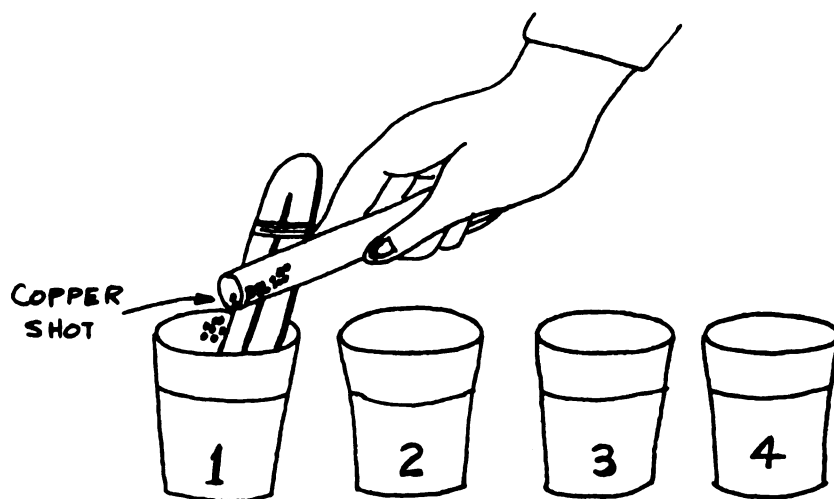
46. Which of these are variables in this experiment?

- | | |
|-----------------|-------------------------|
| 1. temperature | 3. weight of shot |
| 2. kind of shot | 4. time over the burner |

47. What is the energy source?

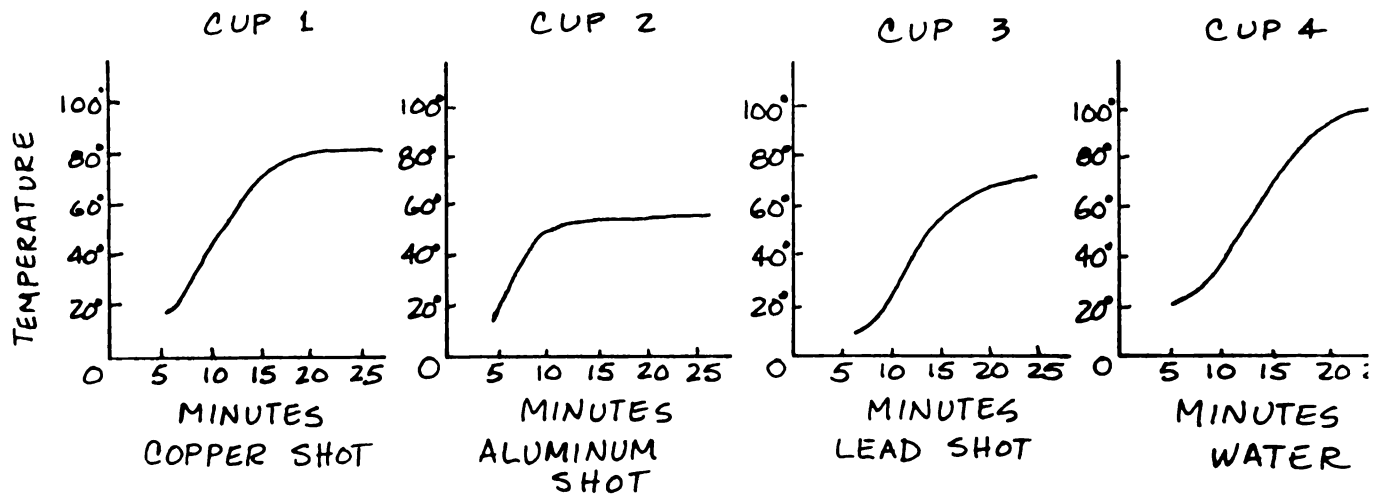
1. the test tubes
2. the boiling water
3. the beaker
4. the thermometer

After the test tubes had been heated for 5 minutes, they were removed from the beaker and their contents poured into separate cups which each contained 100 ml of water at room temperature.



48. What is the heat energy source here?
1. The 100 ml of water in the cups
 2. The material inside each test tube
 3. The cups themselves
 4. The person's hand

The graphs below show the following results of the four cups from the previous pages:



49. Which cup shows the greatest increase in temperature?

1. Cup 1 2. Cup 2 3. Cup 3 4. Cup 4

50. Which material was the best heat source?

1. Copper shot
2. Aluminum shot
3. Lead shot
4. Water

51. Which material gave up its heat the fastest?

1. Copper shot
2. Aluminum shot
3. Lead shot
4. Water

52. This experiment shows that:

1. The faster a substance gives up its heat the more heat it gives up.
2. The faster a substance gives up its heat the less heat it gives up.
3. The slower a substance gives up its heat the more heat it gives up.
4. The slower a substance gives up its heat the less heat it gives up.

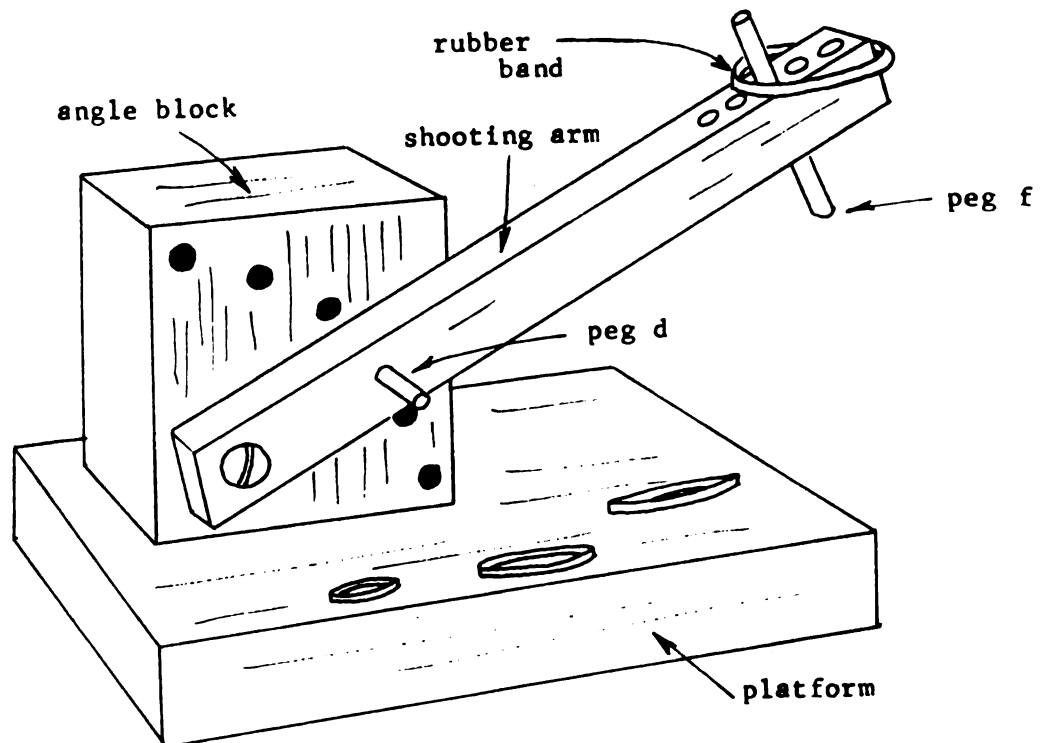
* * * * *

Items 53-61 refer to the following explanation and diagram:

Students in science class were having a contest to determine how far a rubber band will travel when shot with the shooting machine pictured below. The machine contains a number of variables which may affect the distance that the rubber bands would travel.

The students were given 3 rubber bands--1 short, 1 medium, and 1 long. Before doing any experiments with the machine most students felt that the longest rubber band would always travel further than the others. Study the entire diagram and locate the following parts:

- | | |
|-----------------|----------------|
| a. platform | d. peg d |
| b. angle block | e. rubber band |
| c. shooting arm | f. peg f |



For Questions 53-57, decide which of the following applies to each statement, and write in the number of the statement you choose.

- KEY:
1. Something that puzzled the students (problem)
 2. Something that the students thought to be true (assumption)
 3. Something that students could determine by looking only at the system (observation)

53. The length of a rubber band determines how far it will travel

54. The class wanted to find out which rubber band would travel the farthest

55. The machine contains a number of variables which affect the distance that rubber bands travel.

56. The shooting arm can be placed in 6 possible positions

57. Three different rubber bands were used in this experiment

58. The energy receiver in this system would be

- | | |
|--------------------|-----------------|
| 1. the rubber band | 3. the platform |
| 2. the angle block | 4. the peg d |

59. The energy source would be

- | | |
|------------------------------|--------------------|
| 1. the stretched rubber band | 3. the platform |
| 2. peg f | 4. the angle block |

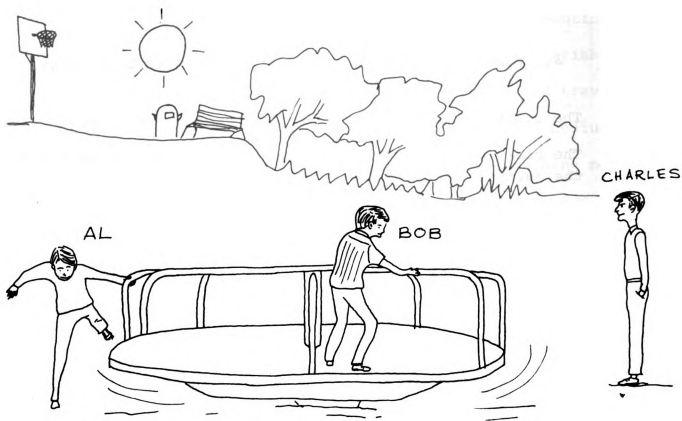
60. In this experiment, what variable would you measure?

1. distance the rubber band travels
2. the angle of the shooting arm
3. length of the rubber band
4. the location in the room of the platform

61. Which of these variables would have the LEAST effect on the distance travelled by the rubber band?

- | | |
|------------------------------|--|
| 1. size of the rubber band | 3. angle of shooting arm |
| 2. thickness of the platform | 4. distance the rubber band is stretched |

Items 62-67 refer to the following picture and explanation:



Study the picture of the children playing with the playground merry-go-round. Locate the following children: Al, Bob and Charles. Suppose someone at the park records the voices of the children while they are playing. Later on, you try to decide who could have made each of the following statements:

62. "You and I are always the same distance apart."
1. Al 2. Bob 3. Charles 4. Al or Bob
63. "I keep running but my position stays the same."
1. Al 2. Bob 3. Charles 4. Al or Charles
64. "The trees keep moving out of sight and then coming back into view."
1. Al and Charles 3. Al and Bob
2. Bob and Charles 4. Al, Bob, and Charles
65. "My distance to the park bench stays the same."
1. Al 2. Bob 3. Charles 4. Al or Charles
66. The energy source for the merry-go-round would be:
1. Al 2. Bob 3. Charles 4. Al and Charles
67. The evidence of interaction in the picture is:
1. the trees
2. the sun
3. the movement of the merry-go-round
4. the basketball net

* * * * *

Table C1.--Summary data on items for TSIS, Form A, second pilot test.

Item Number	Correct Response	Process & Content Area ^a	Item Difficulty	Item Discrim.
1	3	E, 3.0	9	11
2	2	E, 1.1	2	11
3	4	D, 1.3	45	7
4	4	E, 1.2	84	0
5	3	D, 1.0	24	44
6	1	D, 1.0	56	44
7	2	D, 1.0	54	41
8	2	D, 1.3	60	26
9	2	D, 1.0	41	30
10	3	E, 1.3	58	37
11	3	E, 1.0	39	44
12	1	H, 2.0	40	22
13	2	C, 2.3	54	37
14	2	E, 6.0	13	22
15	2	H, 2.0	39	-11
16	2	H, 2.1	50	26
17	2	E, 3.0	20	30
18	2	E, 3.0	34	44
19	2	D, 3.1	12	22
20	2	E, 3.2	37	51
21	4	C, 3.1	28	52
22	2	D, 1.3	35	45
23	1	E, 3.0	56	41
24	3	F, 3.2	48	40
25	2	D, 3.0	45	56
26	2	E, 3.0	45	59
27	2	E, 3.0	40	59
28	2	E, 3.1	55	56
29	1	E, 3.0	46	67
30	2	G, 3.0	61	22
31	1	D, 2.2	67	37
32	3	D, 2.2	45	56
33	1	D, 2.2	57	62
34	3	D, 2.2	53	26
35	3	D, 2.2	67	56
36	3	F, 3.0	53	37
37	2	E, 3.0	61	26
38	3	E, 3.0	70	19
39	3	E, 3.0	67	11
40	3	G, 3.0	70	19
41	3	E, 3.2	65	11
42	3	E, 6.0	31	22
43	2	H, 2.1	27	37
44	3	E, 6.0	45	51

Table C1.--Continued.

Item Number	Correct Response	Process & Content Area ^a	Item Difficulty	Item Discrim.
45	1	E, 6.0	39	41
46	2	C, 2.3	65	14
47	2	E, 6.2	47	51
48	2	E, 6.2	47	30
49	4	D, 6.2	24	45
50	4	D, 6.0	38	59
51	2	D, 6.2	46	37
52	3	D, 6.3	74	7
53	2	D, 6.3	64	22
54	1	A, 6.3	62	44
55	3	D, 6.3	62	14
56	3	D, 6.3	59	55
57	3	D, 6.3	67	18
58	1	E, 6.2	64	30
59	1	E, 6.1	56	44
60	1	C, 6.3	51	26
61	2	E, 6.3	64	48
62	3	D, 4.0	78	0
63	1	D, 4.0	30	41
64	4	D, 4.0	32	54
65	3	D, 4.0	26	29
66	1	E, 6.0	30	52
67	3	H, 2.1	28	60

^aRefer to Figure C1 for process and content areas.

	(A) Recognizing the problem	(B) Formulating a hypothesis	(C) Identifying and controlling variables	(D) Interpreting experimental data	(E) Inferring	(F) Selecting suitable tests for hypotheses	(G) Predicting	(H) Nonprocess items
1.00 Population				4	2			
1.1 Predator-prey								
1.2 Food chain					1			
1.3 Populations interdependence				3	1			
2.00 Subsystems & Variables								2
2.1 Evidence of interaction								3
2.2 Identifying systems				5				
2.21 Isolating subsystems								
2.22 Identifying variables								
2.23 Designing controlled experiments to determine variable's effects								
2.24 Constructing and interpreting graphs and charts								
2.3 Predicting results of a similar experiment			2					
3.00 Environments				1	10	1	2	
3.1 Environmental factor			1	1	1			
3.2 Range of environmental factors					1	2		
3.3 Optimum range								
4.00 Relative Position & Motion				4				
4.1 Relative motion								
4.2 Rectangular coordinates								
4.3 Polar coordinates								
5.00 Communities								
5.1 Plant eater-animal eater								
5.2 Producers-consumers								
5.3 Decomposition								
5.4 Cyclic path of materials								
6.00 Energy Sources				1	5			
6.1 Energy transfer from energy source to energy receiver					1			
6.2 Variables which affect energy transfer				2	3			
6.3 Predicting amount of energy transfer	1		1	5	1			

Figure C1.--Preliminary test specification grid, Form A, second pilot test.

A TEST OF SCIENCE INQUIRY SKILLS

Preliminary Edition by Joseph W. Riley

DIRECTIONS

1. Remove the answer sheet from the test booklet. Close the booklet after removing the answer sheet.
2. Print your name, last name first, in the space provided in the upper left corner of the answer sheet.
3. In the space next to your name, write your teacher's name.
4. In the spaces under your name, write your birth date, for example, January 4, 1961.
5. In the boxes in the upper right corner print in as much of your name as possible and fill in the correct spaces under your name.
6. Find the box labeled "sex" at the bottom of the page. If you are a boy, fill in the space labeled "M". If you are a girl, fill in the space labeled "F".
7. Now turn the answer sheet so that your name is on the left side of the page.
8. Find Number 1 on the answer sheet, then find Number 8. Notice that the numbers run ACROSS the page, from left to right.
9. Now open your test booklet. Each question in this test is followed by 3 or 4 choices for an answer. Read each question and then decide which is the one best answer.
10. On the answer sheet find the row which has the same number as the question you are answering. In that row blacken the space that has the number which is the same as the one in front of the answer you choose.
11. Make your marks heavy and black, press down hard on your pencil, and go over your marks. If you make a mistake, erase completely the answer you wish to change. Use only the special pencil that has been provided.
12. Look at this sample question:

85. A cow is a:

1. machine 2. plant 3. animal 4. rock

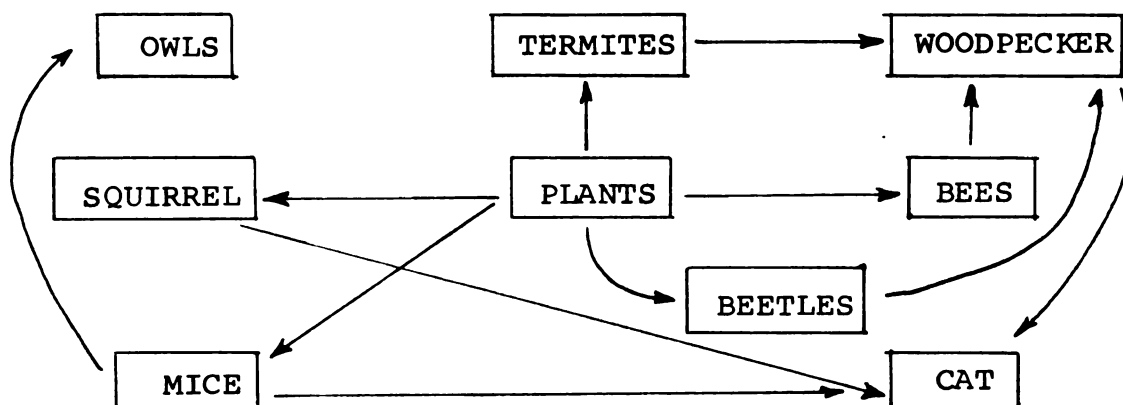
For this question, the answer sheet should look like this:

85. 1 2 ☒ 4

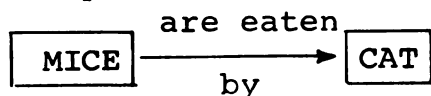
For each item in this examination select the ONE best answer.
Do not mark more than ONE answer space for each item.

* * * *

For items 1-11 use the following picture and example:



KEY: For example --



1. This food web shows that

1. Owls eat cats
2. Mice eat plants, and owls and cats eat mice
3. Bees eat beetles
4. Woodpeckers eat termites, bees, and mice

2. A food chain in the diagram is

1. Plants → beetles → woodpecker
2. Plants → cats → mice
3. Termites → squirrels
4. Plants → squirrels → mice

3. The animal that would be considered a wood decomposer would most likely be

1. owl
2. termite
3. bee
4. cat

4. If the owls were removed from the environment, one would expect an increase in the number of
1. bees 2. plants 3. woodpeckers 4. mice
5. If the owls stayed away forever, probably the number of _____ would increase.
1. bees 2. cats 3. woodpeckers 4. beetles
6. Plants and animals living together as shown in the picture are called a
1. group 2. community 3. population 4. field
7. If more woodpeckers were to move into the area we would expect
1. an increase in the number of termites
2. a decrease in the number of beetles
3. an increase in the number of squirrels
4. a decrease in the number of plants
8. If cats are removed from this area,
1. the number of owls would decrease
2. the number of woodpeckers would decrease
3. the number of mice would increase
4. the number of plants would increase
9. The number of woodpeckers would remain the same if
1. more owls came to the forest
2. all of the beetles died
3. all of the cats left
4. the number of squirrels increased
10. In the situation pictured, you would expect to find the largest number of
1. plants 2. cats 3. owls 4. squirrels

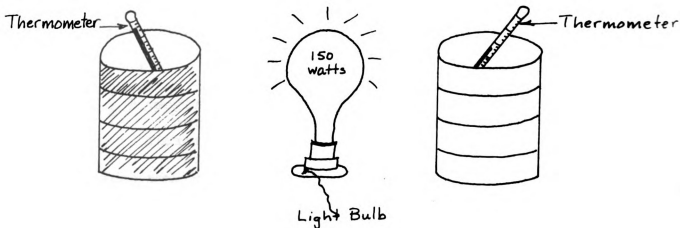
11. The area was sprayed with DDT (a poison) to kill the beetles. A number of owls were found dead due to DDT.

NOTE: These owls don't eat beetles. We can conclude that:

1. The owls got the DDT from the beetles
2. The owls got the DDT from the mice who got it from the plants
3. The owls got the DDT from the plants
4. The owls were not affected by the DDT

* * * * *

Items 12-15 are based on the following drawing and explanation.



In this experiment we have a light bulb and 2 cans that are the same size. Each can is filled with the same amount of water. The outside of one can is painted black and the outside of the other is painted white. These cans are placed at equal distances from a light bulb that is turned on. After 3 hours, the black can is hotter than the white can.

12. The best energy receiver would be

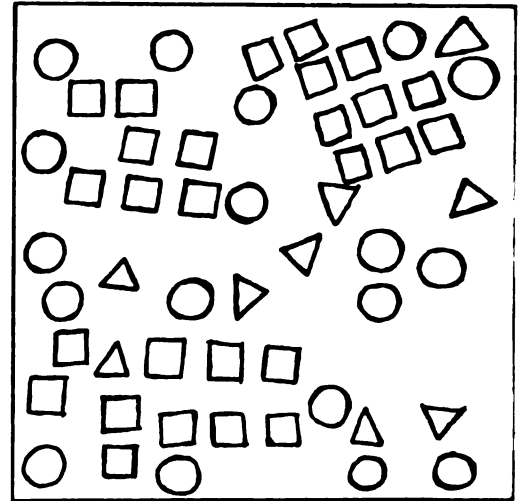
1. The light bulb
2. The black can
3. The white can
4. Both cans are equal

13. The energy source in this experiment would be the
 1. light bulb
 2. black can
 3. white can
 4. thermometer
14. The variable being tested in this experiment is the
 1. amount of water used
 2. size of the can
 3. color of the can
 4. distance between the light and the cans
15. During the daytime in the summer, a house with a white roof should be
 1. warmer than a house with a black roof
 2. cooler than a house with a black roof
 3. the same temperature as a house with a black roof

* * * * *

For items 16-23, use Pictures A and B.

PICTURE A -- represents
several kinds of living
things living in a field.



16. How many different kinds of living things are shown
in the picture?

1. 18 2. 27 3. 54 4. 3

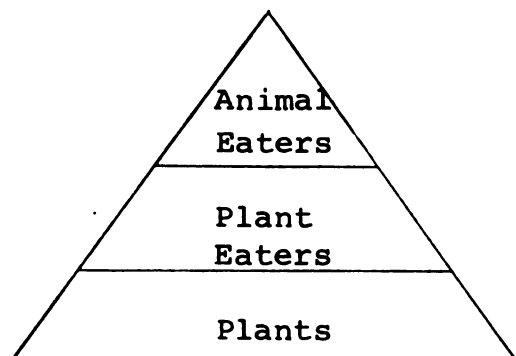
17. All the animals of one kind would be called a

1. population 2. individual. 3. community

18. All of the different kinds of living things living
together would be a

1. population 2. individual 3. community

PICTURE B -- the
triangle represents
the number of living
things of each kind
found in a field.



19. In Picture A, which symbol represents plants?

1. ☐ 2. 3.

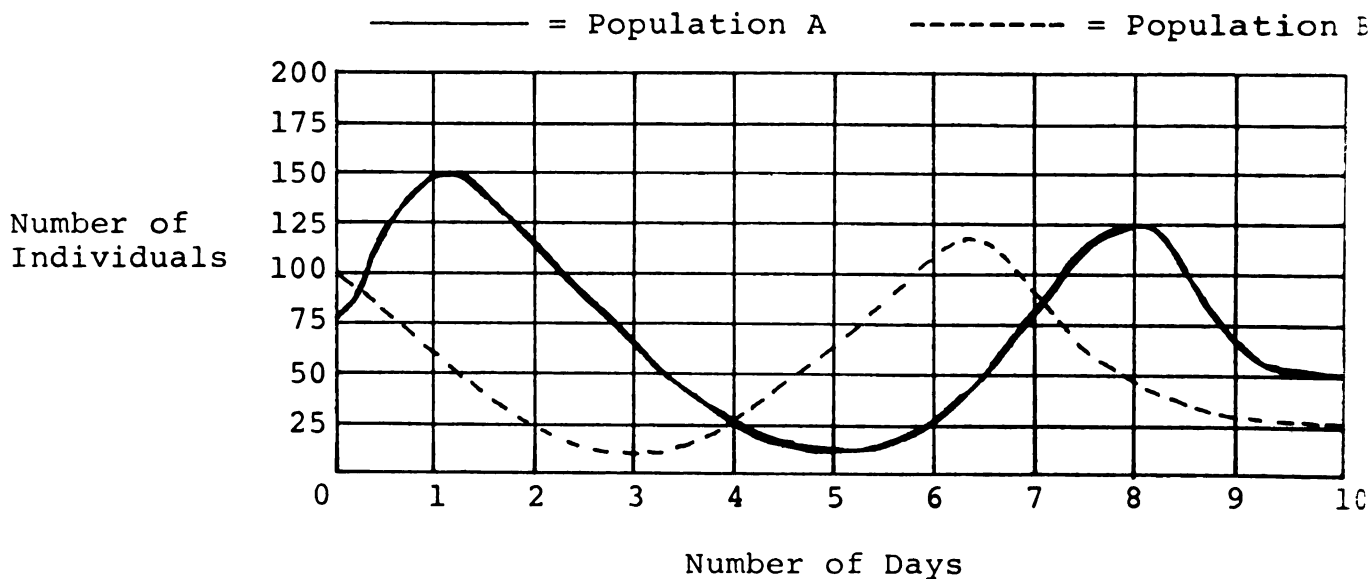
20. In Picture A, which symbol represents plant eaters?

1. ☐ 2. 3.

21. If the number of plant eaters were to decrease, what effect would this have on the number of plants?
 1. They would increase
 2. They would remain the same
 3. They would decrease
22. If all light were kept from this location, what would happen?
 1. All organisms would eventually die
 2. Things would continue as usual
 3. Only the plants would die
23. If some new organisms were added to the group and the number of plant eaters decreased, this could be because the new individuals
 1. were plants
 2. were plant eaters
 3. were animal eaters

* * * * *

Items 24-32 refer to the following explanation and graph. The following questions deal with a community made up of 2 different kinds of water animals. All the animals are living in the same small aquarium. The graph below describes the activities of these two populations.



24. From the graph, how many individuals were in Population A at the beginning of this experiment?
1. 50 2. 75 3. 100 4. 125
25. At the 8th day, what is the number of individuals in Population A?
1. 50 2. 75 3. 100 4. 125
26. At the 8th day, what is the number of individuals in Population B?
1. 50 2. 75 3. 100 4. 125

- * * * * *

34. What conclusion can you draw from this experiment?

1. Cocklebur plants in light produce red flowers
2. The amount of light a cocklebur plant receives determines whether or not it will produce flowers
3. Plastic covers cause cocklebur plants to produce flowers
4. Cocklebur plants must be 3 inches tall to produce flowers

35. If your teacher gave you a cocklebur plant and told you that it had been exposed to light for 18 hours a day, what would you decide, based on what you have learned?

1. The plant is at least 5 inches tall
2. The plant probably needs watering
3. The plant has not produced any flowers
4. The plant has flowers

A second experiment was done using a different kind of plant.

(Black Henbane plants) Some of these plants were exposed to light for 8, 9, 10, 11, 12, 13, or 14 hours each day. The students then looked at the plants to see which ones had flowers. Read the table carefully.

36. How many hours of light do these plants need in order to produce flowers?

- | | |
|----------------------|------------------------------|
| 1. 8 hours of light | 3. 12 hours of light |
| 2. 10 hours of light | 4. This cannot be determined |

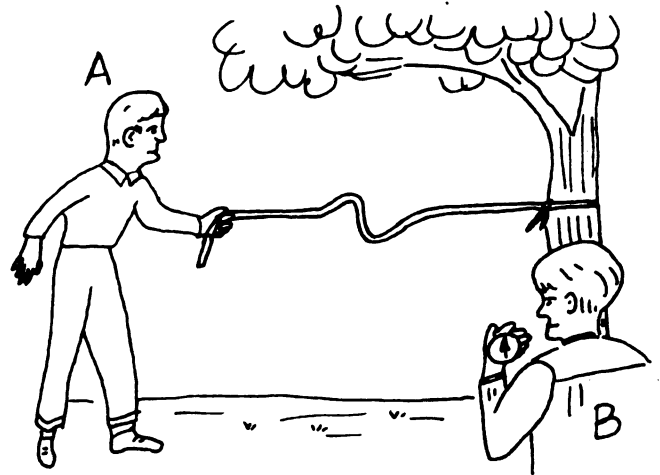
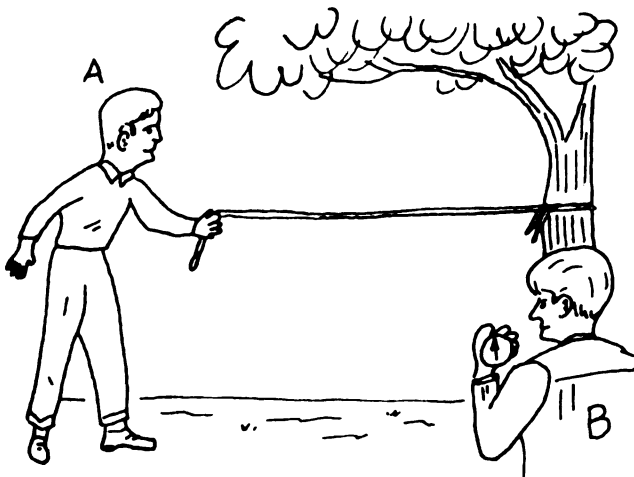
37. If you were to expose a Black Henbane plant to 6 hours of light, what do you predict would happen? The plant would:

- | | | |
|---------------|---------------|------------------------|
| 1. die | 2. need water | 3. have flowers |
| 2. need water | | 4. not produce flowers |

38. If you wanted to grow both cocklebur and black henbane plants together and have both of them produce flowers, how many hours of light would you expose them to?
- | | |
|-------------------|---------------------------|
| 1. 8 to 10 hours | 3. 16 to 18 hours |
| 2. 12 to 14 hours | 4. Not possible to decide |
39. If you wanted to grow these plants together and have neither of them produce flowers, how many hours of light would you expose them to?
- | | |
|-------------------|---------------------------|
| 1. 8 to 10 hours | 3. 16 to 18 hours |
| 2. 12 to 14 hours | 4. Not possible to decide |

* * * * *

Items 40-42 refer to the following pictures:



40. In the pictures above, what is the energy source that caused the change in the rope's pattern?

- | | |
|--------------|-------------|
| 1. The tree | 3. The rope |
| 2. The clock | 4. Boy A |

41. The evidence of interaction in the rope, tree, boy system is the movement of

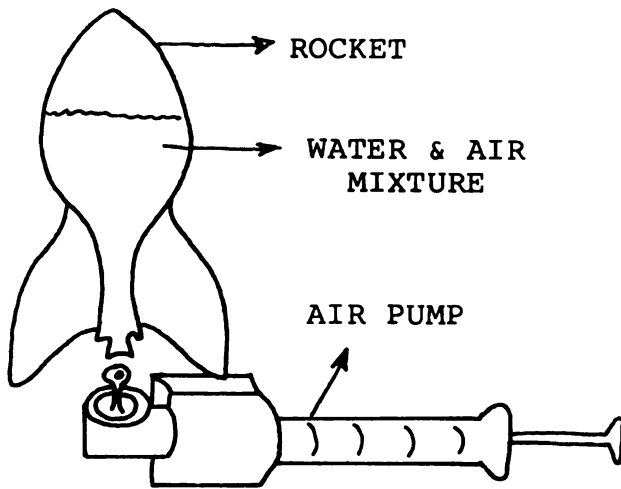
- | | |
|---------------------------|-------------------|
| 1. the hands on the clock | 3. the boy's hand |
| 2. the loop in the rope | 4. the tree |

42. If more energy were used in moving the end of the rope up and down, how would the pictures be different?

1. Boy A would be taller
2. The clock would run faster
3. The size of the loop would be larger
4. The rope would be longer

* * * * *

Items 43-48 refer to the following diagram and explanation:



PARTS:

- I. Air Pump
- II. Rocket
- III. Water and Air Mixture

Perhaps you have seen a water rocket. It consists of the rocket itself and an air pump which fits on the bottom of the rocket. The person using the rocket adds the amount of water he wishes to the rocket, then places the air pump on the bottom of the rocket and pumps air into the rocket. The amount of pressure increases as the number of pumps is increased.

43. What is included in this system?

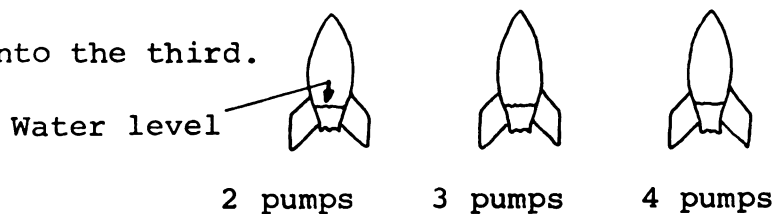
- 1. I only
- 2. I and II
- 3. II and III
- 4. I, II and III

44. What is the energy source in this system?

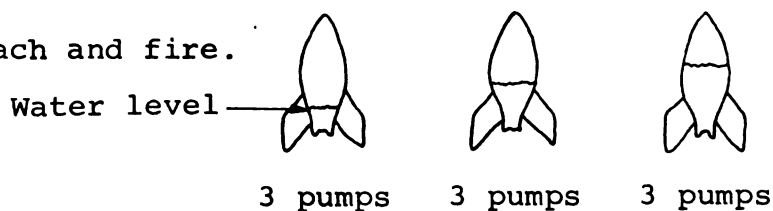
- 1. The rocket
- 2. The water
- 3. The air
- 4. The compressed air and water mixed in the rocket

45. Suppose a person wanted to find out how the number of pumps of air in the rocket affected the height the rocket could rise. Which experiment would be best to determine this?

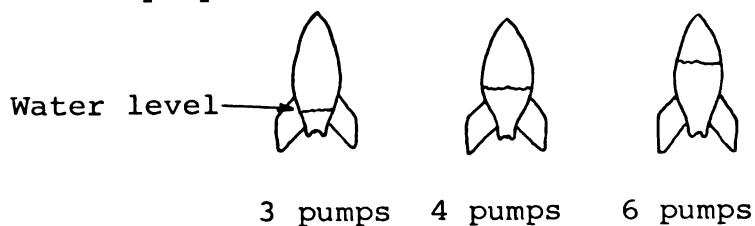
1. Take 3 rockets; place 20 ml of water into the first one, 20 ml of water in the second, and 20 ml of water in the third. Put 2 pumps full of air into the first; then 3 pumps full of air into the second, and 4 pumps full of air into the third.



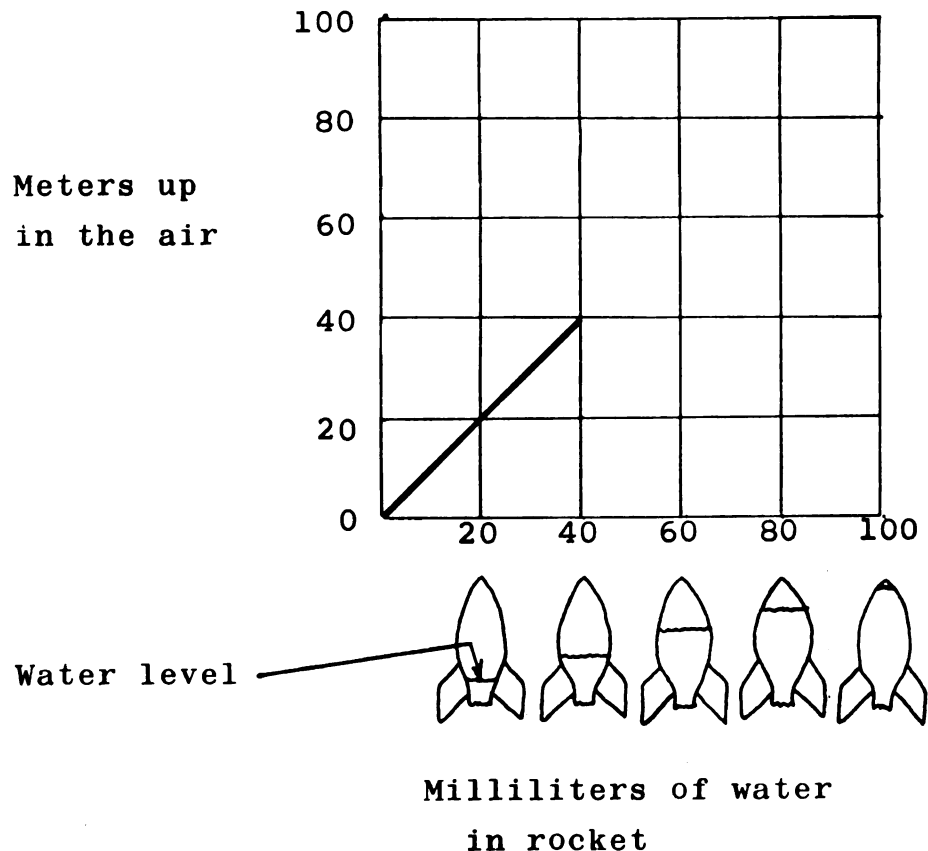
2. Take 3 rockets; place 20 ml of water in one, 40 ml in the second, and 60 ml in the third. Put 3 pumps full of air into each and fire.



3. Take 3 rockets; place 20 ml of water into the first, 40 ml of water into the second, and 60 ml into the third. Put 3 pumps full of air in the first, 4 pumps in the second, and 6 pumps in the third.



Suppose you were given the following information: Some tests were made using rockets with the amount of water varied as shown on the bottom of the graph and the height that the rocket rose on the side of the graph. Each rocket was given the same number of pumps of air.



46. How high would you predict the rocket with 60 ml of water would go?
1. 45 meters
 2. 60 meters
 3. 80 meters
 4. 100 meters

When the rocket with 70 ml of water in it was fired, it was found to go up 45 meters. When the rocket with 80 ml of water in it was fired, it was found to go up 40 meters.

47. Knowing this, how high would you predict the rocket with 100 ml of water in it would go?

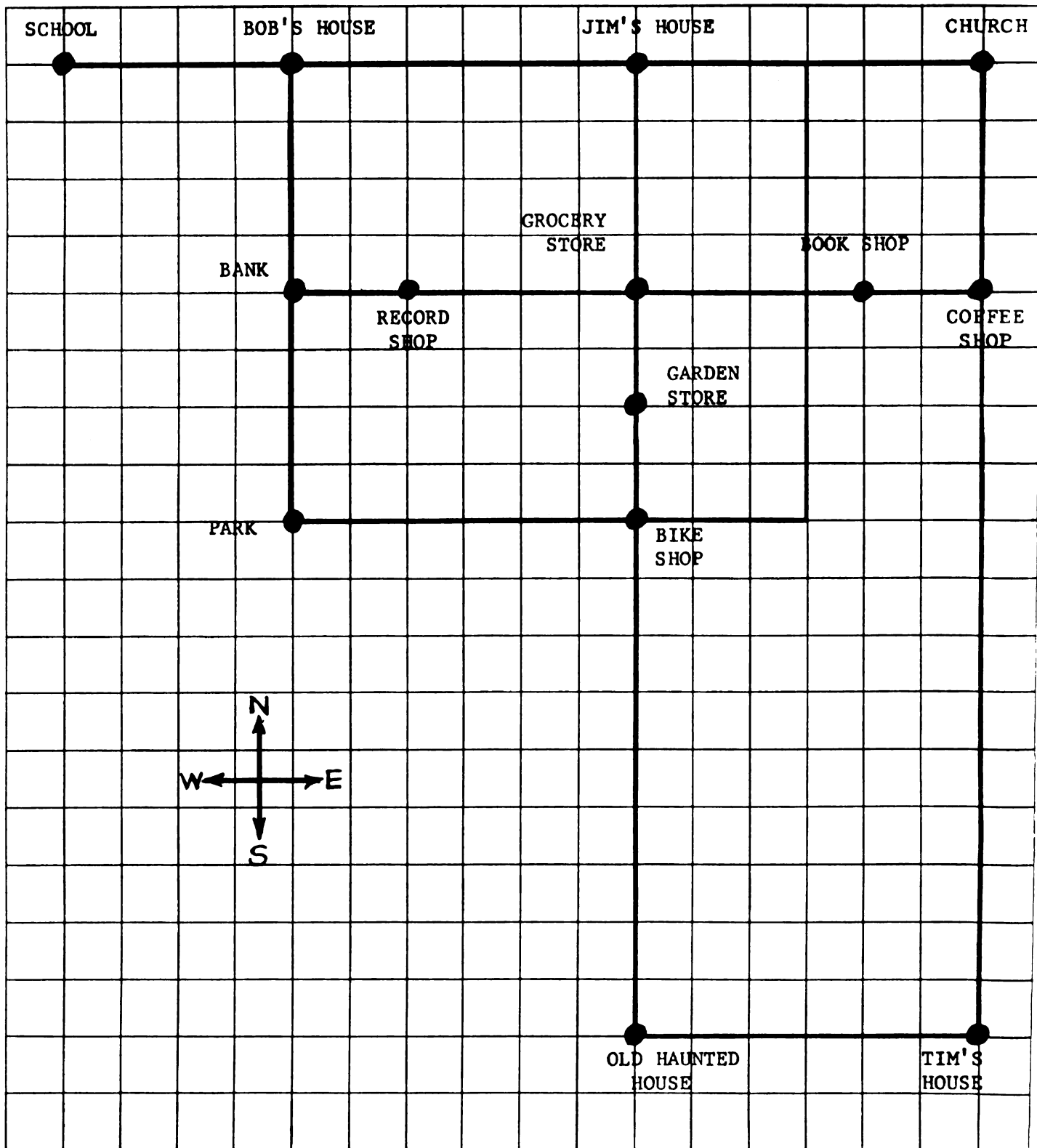
1. 35 meters
2. 40 meters
3. 60 meters
4. 80 meters

48. Based on the evidence that has been presented, what seems to affect the height of the rocket?

1. The amount of air in the rocket
2. The amount of water in the rocket only
3. The combination of air and water in the rocket
4. The strength of the person doing the pumping

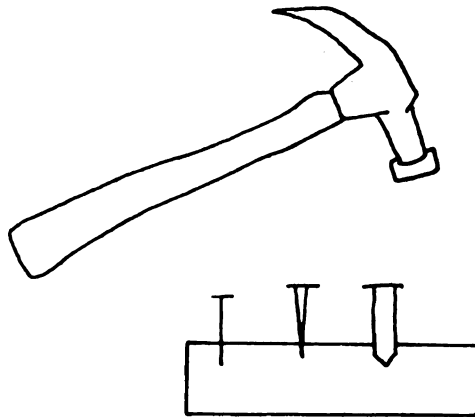
* * * * *

Items 49-54 refer to the following map. On this map, each space between 2 lines represents a distance of 1 unit. The compass in the bottom left corner tells the directions.

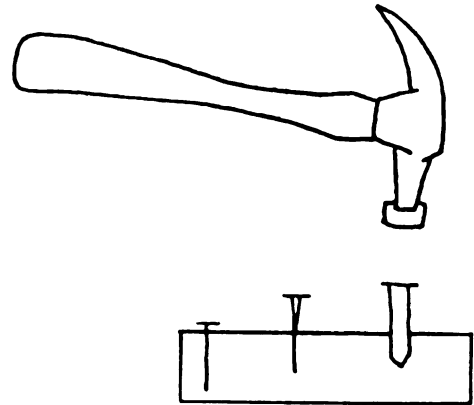


49. How many units is it from the school to Jim's house?
1. 6 units 2. 8 units 3. 10 units 4. 12 units
50. To get to school from Bob's house you would have to travel in which direction?
1. North 2. South 3. East 4. West
51. If you were at the Grocery Store and went south 4 units, where would you be?
1. Bike Shop 3. Book Shop
2. Old Haunted House 4. Jim's House
52. Suppose Jim left his house for a walk and went east 6 units, then south 4 units and then west 2 units. Where would he now be?
1. Record Shop 3. Book Shop
2. Grocery store 4. Garden Store
53. If Bob was at the Bike Shop and had to purchase bread for his mother on the way home, which route would be the best for him to take?
1. West 6 units; North 8 units
2. North 4 units; West 6 units; North 4 units
3. East 4 units; North 8 units; West 9 units
4. North 4 units; East 6 units; North 4 units
54. If Pete told you he left school and followed the following path from school: A) East 10 units; B) South 4 units; C) West 6 units; D) South 4 units; E) East 6 units
Where is he now?
1. Garden Shop 2. Tim's House 3. Bike Shop 4. Book Shop

Items 55-58 refer to the following pictures and explanations.



PICTURE A
BEFORE



PICTURE B
AFTER

Picture A shows a board with 3 nails before the start of an experiment. Picture B shows the same board after each nail was hit by the same person using the same hammer, three times.

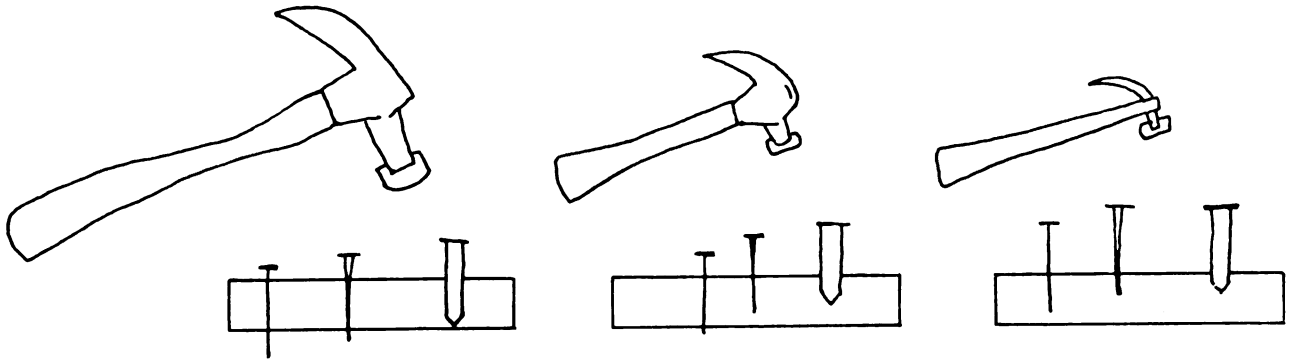
55. The energy source would be

- | | |
|---------------|------------------------|
| 1. The hammer | 3. The wood |
| 2. The nail | 4. The small nail only |

56. Which variable accounted for the results shown in Picture B?

1. The size of the hammer
2. The thickness of the nail
3. The thickness of the wood
4. The number of times each nail was hit

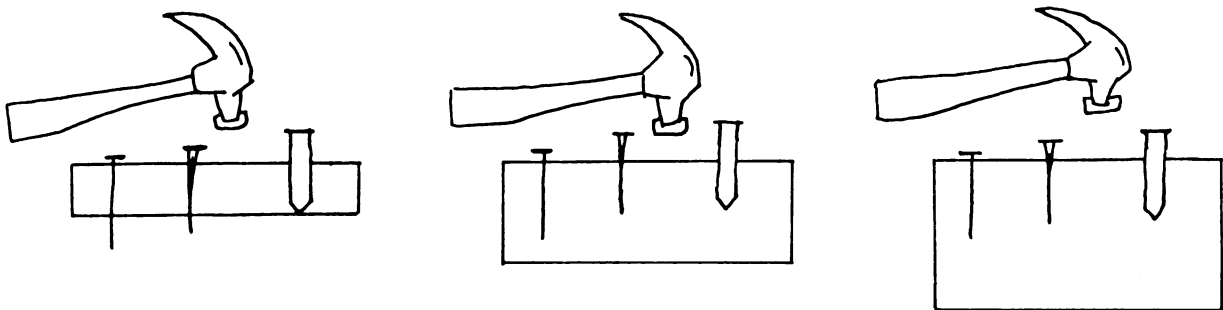
A student, using 3 different hammers to hit the nails 3 times each, had the following results shown in these pictures:



57. Which variable most likely accounted for the results shown in these pictures?

1. The hammers that were used
2. The size of the nails
3. The thickness of the wood
4. The number of times each nail was hit

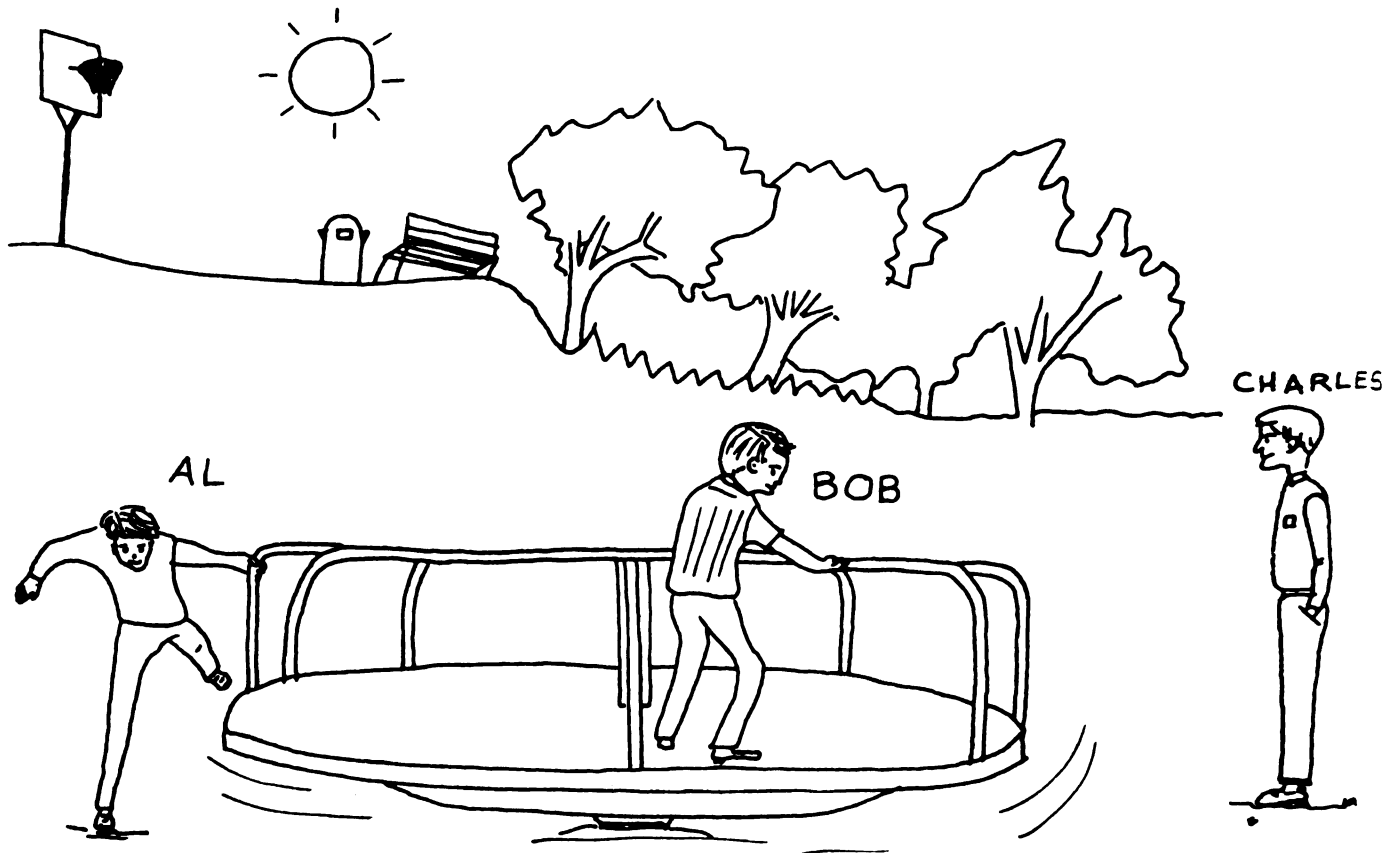
Another experiment was tried with a student using the same hammer and hitting each nail 3 times. The results are shown below:



58. Which variable probably accounts for the results shown in these pictures?

1. The size of the hammer
2. The size of the nails
3. The thickness of the wood
4. The number of times each nail was hit

Items 59-65 refer to the following diagram and explanation:



Study the diagram of the children playing with the playground merry-go-round. Locate the following children: Al, Bob, and Charles. Suppose someone at the park records the voices of the children while they are playing. Later on, you try to decide who could have made each of the following statements:

59. "You and I are always the same distance apart."

1. Al 2. Bob 3. Charles 4. Al or Bob

60. "I keep running but my position stays the same."

1. Al 2. Bob 3. Charles 4. Al or Charles

61. "The trees keep moving out of sight and then coming back into view."
1. Al 2. Bob 3. Charles 4. Al or Bob
62. "My distance to the park bench stays about the same."
1. Al 2. Bob 3. Charles 4. Al or Charles
63. "If you keep moving away from me, you are going to fall off the edge."
1. Al 2. Bob 3. Charles 4. Al or Charles
64. The energy source for the merry-go-round would be:
1. Al 2. Bob 3. Charles 4. Al and Charles
65. The evidence of interaction in the picture is
1. The trees
2. The sun
3. The movement of the merry-go-round
4. The basketball net

* * * * *

Items 66 - 69 refer to the following picture and explanation:



Study the above picture carefully. Locate the following children:

A = Alice

B = Betty

C = Cathy

D = David

Suppose someone records the voices of the children while they are playing. Later on, you try to decide who could have made each of the following statements:

66. "The people on the swings keep going back and forth."

1. Alice and David
2. Betty and Cathy
3. Cathy and Alice
4. David and Betty

67. "The people on the swings keep going away from me and then towards me."

1. Alice
2. Betty
3. Cathy
4. David

68. All the children are moving. Which children are moving together?

1. Alice and Betty
2. Betty and Cathy
3. Cathy and David
4. Alice and David

69. "The ground keeps getting closer and then farther away."

1. Alice and Betty
2. Betty and Cathy
3. Cathy and David
4. Alice and David

Table C2.--Summary data on items for TSIS, Form B, second pilot test.

Item Number	Correct Response	Process & Content Area ^a	Item Difficulty	Item Discrim.
1	2	E, 1.2	29	34
2	1	H, 1.3	40	56
3	2	H, 5.3	24	44
4	4	G, 1.1	23	48
5	2	G, 1.1	60	44
6	2	H, 5.0	53	48
7	2	G, 1.1	45	70
8	3	G, 1.1	19	39
9	1	G, 1.3	66	4
10	1	G, 5.2	62	-13
11	3	E, 5.4	45	44
12	2	E, 6.1	34	53
13	1	E, 6.0	22	34
14	3	C, 6.2	24	57
15	2	E, 6.1	29	48
16	4	D, 1.0	34	48
17	1	H, 1.0	66	-17
18	3	H, 5.0	47	44
19	1	D, 5.1	66	17
20	3	D, 5.1	60	22
21	1	G, 3.3	33	57
22	1	G, 1.3	43	35
23	3	E, 5.1	44	44
24	2	D, 1.0	31	57
25	4	D, 1.0	35	70
26	1	D, 1.0	35	7
27	3	D, 1.0	13	39
28	4	D, 1.0	44	40
29	2	D, 1.0	41	61
30	3	D, 1.0	65	48
31	1	G, 1.2	41	48
32	1	E, 2.1	50	26
33	1	C, 3.1	28	48
34	2	E, 3.2	43	44
35	3	D, 3.1	65	0
36	3	D, 3.2	45	44
37	4	G, 3.3	58	48
38	2	C, 3.2	41	48
39	4	G, 6.1	83	21
40	4	E, 6.0	27	53
41	2	H, 2.1	76	30

Table C2.--Continued.

Item Number	Correct Response	Process & Content Area ^a	Item Difficulty	Item Discrim.
42	3	G, 6.2	40	61
43	4	H, 2.0	43	48
44	4	E, 6.0	56	53
45	1	F, 6.2	67	44
46	2	G, 6.1	56	61
47	1	G, 6.1	62	65
48	3	E, 6.3	47	18
49	3	D, 4.0	27	44
50	4	D, 4.0	35	35
51	1	D, 4.0	29	48
52	3	D, 4.0	31	51
53	2	D, 4.0	65	56
54	3	D, 6.0	41	70
55	1	E, 6.0	31	78
56	2	C, 2.21	55	44
57	1	C, 2.21	44	35
58	3	C, 2.21	47	52
59	4	D, 4.0	34	50
60	1	D, 4.0	34	57

^aRefer to Figure C2 for process and content areas.

	(A) Recognizing the problem	(B) Formulating a hypothesis	(C) Identifying and controlling variables	(D) Interpreting experimental data	(E) Inferring	(F) Selecting suitable tests for hypotheses	(G) Predicting	(H) Nonprocess items
1.00 Population				8				1
1.1 Predator-prey							4	
1.2 Food chain					1		1	
1.3 Populations interdependence							2	
2.00 Subsystems & Variables								1
2.1 Evidence of interaction					1			2
2.2 Identifying systems								
2.21 Isolating subsystems			3					
2.22 Identifying variables								
2.23 Designing controlled experiments to determine variable's effects								
2.24 Constructing and interpreting graphs and charts								
2.3 Predicting results of a similar experiment								
3.00 Environments								
3.1 Environmental factor			1	1				
3.2 Range of environmental factors			1	1	1			
3.3 Optimum range							2	
4.00 Relative Position & Motion				14				
4.1 Relative motion								
4.2 Rectangular coordinates								
4.3 Polar coordinates								
5.00 Communities								2
5.1 Plant eater-animal eater				2	1			
5.2 Producers-consumers							1	
5.3 Decomposition								1
5.4 Cyclic path of materials					1			
6.00 Energy Sources				1	5			
6.1 Energy transfer from energy source to energy receiver					2		3	
6.2 Variables which affect energy transfer			1			1	1	
6.3 Predicting amount of energy transfer					1			

Figure C2.--Preliminary test specification grid, Form B,
second pilot test.

APPENDIX D

OVERVIEW AND INSTRUCTIONS GIVEN TO PARTICIPATING PRINCIPALS AND TEACHERS

OVERVIEW AND INSTRUCTIONS GIVEN TO
PARTICIPATING PRINCIPALS AND TEACHERS

May 1, 1972

In talking with school administrators and teachers and in reading the literature, I have detected a real need for more information concerning the relative strengths and weaknesses of the new lab-centered elementary science programs as compared with the textbook-centered programs. Many elementary schools are under increased pressure to upgrade and/or change the science curriculum. However, school administrators and teachers find themselves in a position of making important and expensive decisions without any real information as to what to expect of their students if they adopt one of the new programs.

It has become increasingly evident that there is a need for definitive information on what to expect of children who have had the new lab-centered elementary science programs. As far as I can detect from the literature, sales brochures, etc. for these programs, they are good and children learn more with them; however, little is presented in the way of concrete evidence to substantiate these claims.

I have designed a research dissertation hopefully to make a small contribution in clearing up this lack of information. My dissertation is designed to compare 5th grade

who have been in the Science Curriculum Improvement Study (SCIS) for five years with similar students who have not used the SCIS program.

I have selected three areas in which to compare these children: (1) Science Achievement, (2) Science Inquiry Skills (Processes), and (3) Attitudes Towards Science and Scientists. These areas will be measured using paper and pencil tests which can easily be administered by individual classroom teachers.

Science Achievement will be measured using the Elementary Science Test of the Sequential Test of Educational Progress (STEP II). This is a nationally normed instrument developed by the Educational Testing Service. It consists of 50 items and requires 40 minutes to administer. The test measures a student's knowledge and understanding of fundamental concepts and processes of science.

Science Inquiry Skills (Processes) will be measured using a test which I have developed in connection with this study. Since there is no test available to measure the processes as stressed in the SCIS program, I had to write one. The 50 items in my test were selected from many items which I developed. They were field-tested by over 1,000 students throughout the state earlier this year. The test measures the following processes: (1) Identifying and Controlling Variables, (2) Interpreting Data, (3) Predicting, and (4) Inferring. This test takes approximately one hour to administer.

Attitudes Towards Science and Scientists will be measured using an instrument developed by Dr. LaMoine Motz. It presents a series of statements about science and scientists which the students are to react to by either agreeing, disagreeing, or remaining undecided. This instrument takes about 20-30 minutes to administer.

General Procedure

- 1) I will provide, deliver, and pick up all materials needed for this study.
- 2) Student answer sheets will be machine scored and upon completion of the study, I will send each school a complete set of results.
- 3) Instructions for administering each individual test will be provided to each teacher.
- 4) Enough copies of each test will be provided so each student will have his own copy of each test.

Schedule

- May 8: Delivery of test booklets, answer sheets, and teacher instructions for: 1) Attitudes Towards Science and Scientists Inventory (SASAI), and 2) Test of Science Inquiry Skills (TSIS).
- May 15: Delivery of test booklets, answer sheets and teacher instructions for the Sequential Test of Educational Progress (STEP).
- May 22: Pick up completed answer sheets and tests for all three tests.*

*This means that teachers have from May 8-19 to administer the TSIS and SASAI tests, and from May 15-19 only to administer the STEP test.

If any problems or questions should arise, please call

Dr. Glenn Berkheimer
Science and Mathematics Teaching Center
Michigan State University
Phone 355-1725

He will relay all messages to me and I will return your call
immediately.

TEACHER'S INSTRUCTIONS
A TEST OF SCIENCE INQUIRY SKILLS

MATERIALS

Test booklet (white & thick)
Answer sheet
Pencil and eraser

GENERAL DIRECTIONS

There is no time limit on this test. My interest is that each student do the best that he possibly can on each item. It is my suggestion that the test be given in at least two testing sessions. However, you may decide that more sessions are appropriate to obtain a more accurate sample of the child's best effort. After the initial testing period in which the answer sheets will be filled out, it is not necessary that all students take the test simultaneously. They can complete the test independently during their free time if this is more appropriate for your classroom situation.

Most sets of questions contain three or four items. Students should be encouraged to try each set. Some of the easiest items are near the end of the test.

Feel free to pronounce difficult words for the students. My goal, as much as possible, is to have this function as a science test, not a reading test.

Make sure that the pupils understand that most of the items present a situation by a picture or diagram and

then a series of questions that can be answered based on the information contained in the diagram. Encourage students to refer to the appropriate diagram as they answer the questions.

1. Distribute materials.
2. Fill in and discuss the proper method of filling out the answer sheet. Instructions are included on the cover of the student manual. Read these instructions along with the children and see that they fill in the information as requested. This will enable me to return your students' results to you. Also, make sure the students understand that the answer columns run ACROSS the page, NOT up and down.
3. Permit the students to begin the test.
4. When the students stop, have them place the answer sheet in the page where they are working so that their name extends beyond the test booklet for easy identification.
5. Second test session: Make certain the students get their own tests again and that they start where they left off after the first session.
6. Collect tests and answer sheets after the students have finished the test. All materials will be picked up on May 22, 1972.

TEACHER'S INSTRUCTIONS

SEQUENTIAL TEST OF EDUCATIONAL PROGRESS (STEP)

MATERIALS

Test booklet
Answer sheet
Pencil and eraser

GENERAL DIRECTIONS

1. Hand out materials.
2. Fill in and discuss proper method of marking the answer sheet.
3. Read the directions on the front cover aloud to the students as they follow along.
4. Open the test book to the page where the directions are and have the students fold their books back so that only the directions and examples show.
5. Read these directions aloud to the students while they follow along. Point out that the correct answer in the example is the one marked with a black mark.
6. Permit the students to begin.
7. This is a 40-minute test.
8. Collect all test books and answer sheets at the end of the 40-minute testing period. These materials will be picked up on May 22, 1972.

TEACHER'S INSTRUCTIONS
SCIENCE AND SCIENTISTS ATTITUDE INVENTORY
(SASAI)

MATERIALS

Test booklet
Pencil and eraser

GENERAL DIRECTIONS

1. Students will fill out the information requested on the front cover of the survey: name, school, sex, and age.
2. The directions are to be read aloud to the class by the teacher.
3. Make certain that students understand that they are to underline their choice in the test booklet.
There is no answer sheet.
4. Point out to the students that the test starts on the back of the front cover.
5. Feel free to pronounce difficult words for students.
6. The test should be given in one sitting.
7. When the students have finished, collect tests and answer sheets. These materials will be picked up on May 22, 1972.

APPENDIX E

LIST OF PARTICIPATING SCHOOLS

ADDRESSES OF SCHOOLS PARTICIPATING
IN THE STUDY

DeWitt Public Schools

1. Fuerstenau Elementary School
205 W. Washington Avenue
DeWitt, Michigan 48820

Grand Ledge Public Schools

1. Delta Center Elementary School
305 S. Canal Road
Lansing, Michigan 48917
2. Delta Mills Elementary School
6816 Delta River Drive
Lansing, Michigan
3. Holbrook Elementary School
615 Jones Street
Grand Ledge, Michigan

Haslett Public Schools

1. Murphy Elementary School
1875 Lake Lansing Road
Haslett, Michigan 48840

Perry Public Schools

1. Perry Elementary School
412 N. Watkins Road
Perry, Michigan 48872

MICHIGAN STATE UNIV. LIBRARIES



31293102354465