Wi

processes of

that test th

Th

items which

use the proc

data. The p

mining which

used by most

vidual test

formance mea

study were t

the process

7 and 9), an

from the pro

9, 12 and 13

Each student

tency Measur

ABSTRACT

THE DEVELOPMENT OF ITEMS WHICH ASSESS THE
PROCESSES OF CONTROLLING VARIABLES
AND INTERPRETING DATA

By

Richard Wayne Robison

With the increasing emphases in instruction in the processes of science, there is a need to develop test items that test these processes in an efficient way.

This study is an attempt to develop objective test items which will assess seventh grade students' ability to use the processes of controlling variables and interpreting data. The procedure followed in constructing and determining which items could be used was not the same as that used by most standardized test writers. The success of individual test items was determined by using independent performance measures. The external criteria used in this study were two of the Individual Competency Measures for the process of controlling variables (Controlling Variables 7 and 9), and three of the Individual Competency Measures from the process of interpreting data (Interpreting Data 9, 12 and 13), all from Part 6 of Science--A Process Approach. Each student was tested individually on the selected Competency Measures and was scored as to whether he or she met the

behavioral

individual

W. Fyffe, a

Processes".

variables,

defining ope

interest to

sample of 5

School in Ch

It

index of ite

the item res

relation to

his showed t

of controlli

for the proc

discriminati

level of acc

An

lation coeff

test and the

crimination

correlation

significant

group scores

of controlli

between the

G37810

behavioral objectives specified in the Measures. After the individual testing, the author, in cooperation with Darrel W. Fyffe, assembled a 79 item group test, "A Group of Four Processes". The test covered the processes of controlling variables, interpreting data, formulating hypotheses and defining operationally. The latter two processes were of interest to Fyffe. The group test was administered to a sample of 56 students who had attended the Kinawa Middle School in Okemos, Michigan, during the sixth grade.

Item analysis data were obtained which reported the index of item discrimination, index of item difficulty, and the item response pattern. The analysis of items was in relation to the external criterion measure. The item analysis showed that 13 of the original 18 items for the process of controlling variables, and 22 of the original 25 items for the process of interpreting data had an index of item discrimination of .20 or greater (this was the predetermined level of acceptance).

Another analysis was performed to show the correlation coefficients between the processes on the individual test and the processes on the group items with high discrimination (.20 or greater). The Pearson Product moment correlation coefficient showed a significant correlation (significant at the .001 level) between the individual and group scores for each of the two processes. For the process of controlling variables, the correlation coefficient between the group items and the individual measure was .70, while the

was .66.

In

sible to d

cantly cor

processes

while the coefficient for the process of interpreting data was .66.

The results of this study indicate that it is possible to develop objective test items which are significantly correlated with students' ability to perform the processes of controlling variables and interpreting data.

THE DEVELOPMENT OF ITEMS WHICH ASSESS THE PROCESSES OF CONTROLLING VARIABLES AND INTERPRETING DATA

By

Richard Wayne Robison

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Secondary Education and Curriculum
College of Education

ACKNOWLEDGMENTS

In a project of this magnitude, there are always many people who help. I cannot possibly thank all of them in this short space, but there are several who must be mentioned.

First, there is one person without whom this project would never have been possible. I refer to my advisor, Dr. Richard J. McLeod. It was he who first suggested this topic to me and who has been a constant source of guidance throughout the long task.

The other members of my committee, Dr. Julian R. Brandou, Dr. Dale Alam and Dr. Norman T. Bell, have also been of invaluable assistance. Their scholarly criticisms have added considerable to the quality of my work.

A special note of thanks to Darrel W. Fyffe, my partner in research and two of his committee members, Dr. Glenn C. Berkheimer and Dr. Robert L. Ebel.

I must also express my thanks to the administration and staff of the Kinawa Middle School who allowed me to do this study.

Finally--my wife and our seven children who had to put up with my weird working hours for years and who truly pitched in to help with all the hard work--it truly must be love.

TABLE OF CONTENTS

CHAPTER

LIST OF TABLES

Table Page

LIST OF FIGURES

Appendi

A.

B.

C.

LIST OF APPENDICES

Backr

scien

never

ment

There

valid

studen

tive

and i

relat

and v

Yearb

are s

These

tific

these

and u

One w

ship
crite

CHAPTER I

THE PROBLEM

Background

The new curriculum materials for elementary school science emphasize the processes of science to an extent never before realized. Yet, there is very little development work underway on the tools for their evaluation. There is an urgent need for a time efficient, reliable, and valid means to evaluate the process-skills obtained by students. The purpose of this study is to develop objective test items over the processes of controlling variables and interpreting data that will give evidence of criterion-related validity[1] while still being time efficient, reliable, and valid.

The goals of teaching science as listed in the 31st yearbook of the National Society for the Study of Education are similar to those listed in the Society's 46th yearbook. These goals were to develop an understanding of the scientific principles and an ability to apply them. Included in these was the development of children's ability to acquire and use scientific attitudes and skills in problem solving. One would find it difficult to quarrel with these statements

Criterion-related validity pertains to a relationship between the scores on test items and an independent criterion measure that defines what is being measured.

as pa

in th

said

recog

then

child

only

was s

who w

been

schoo

done

proje

prac

Year

inte

to d

gain

Elem
Co.,

tion
Nov.

as part of educational objectives. P. Hurd and J. Gallagher, in their book, New Directions in Elementary Science Teaching, said that these "objectives have meaning only as they are recognizable in the changed behavior of children."[2] They then point out that "These objectives were not achieved by children, and the way elementary science was taught could only insure that these goals would not be attained."[3] There was a pressing need for our schools to develop students who were adept in those processes about which so much had been written.

The concern about science teaching in the elementary school reached a point in 1960 where "something had to be done."[4] By 1961 several projects were underway. These projects have included, as part of their goals, the seldom practiced aims that were written in the 31st and 46th NSSE Yearbooks. In general, these projects have included the intellectual skills that would be needed if the student were to discover, organize, and apply the information that he gains from his science experience.

W.W. Welch[5] points out that when new projects are

---

[2]Hurd, Paul and James Gallagher, New Directions in Elementary Science Teaching, California: Wadsworth Publishing Co., Inc., 1969, p. 28.

[3]Ibid., p. 28.

[4]Ibid., p. 29.

[5]W.W. Welch, "Curricular Decisions: How can Evaluation Assist Science Teachers?", The Science Teacher, Vol. 35, Nov. 1968, pp. 22-25.

being
These
effec
Welon
suppo
answe
materi
resou
has al
summe
train

devel
for t
been
two p
newer
often
ioral
these

Measu
objec
A Pro
Educa

being developed, the authors need to answer two questions. These questions are: (1) How can these programs be taught effectively? and (2) What are the methods of evaluation? Welch also points out that most of the newer federally supported programs have provided financial assistance in answering the first question. These programs have prepared materials such as teachers' guides, laboratory manuals, resource books and teacher training films. The government has also financially aided the development of special summer institutes in an effort to provide assistance in training teachers to teach the new programs.

The answer to the second question, the problem of developing methods of evaluation, has been another matter for the newer curriculum projects. It has, by no means, been given the attention that it deserves. Welch points out two problems in developing evaluative materials for the newer science curricula. "First, curriculum developers are often hesitant to state their objectives explicity in behavioral terms. Second, they are even more reluctant to have these objectives evaluated."[6]

In the development of behavioral objectives and measures to determine the extent that students meet these objectives, one program stands out. This program, Science--A Process Approach, developed by the Commission on Science Education of the American Association for the Advancement of

---

[6] Ibid.

Science (AAAS), has provided the materials to help teachers teach the processes to be determined. It has provided the teacher with explicitly stated objectives that state, in behavioral terms, what they expect each student to be able to do at the end of each lesson.

Science--A Process Approach has also provided a means for the teacher to evaluate whether the student has achieved the objectives. To evaluate a student, the teacher constructs a situation in which the student is given materials with which to work. The teacher then asks questions which involve the use of the process being tested and observes the individual as he manipulates the materials in an attempt to answer the questions. Since the teacher must infer from the student's actions, his ability to exhibit the acceptable behaviors, it must be administered by someone who is familiar with the specific process involved. There is a need for an objective test that could be administered by anyone, even one unfamiliar with the behaviors that define the process tested.

Another problem associated with evaluating the process competency of a student is that it requires the evaluator to work with one student at a time. A group test would be more efficient.

The amount of equipment that the evaluator must have at hand for evaluation creates still another problem. There is a need to develop a test over the processes of science that will minimize the amount of equipment needed for administration.

Although the above points (objectivity, efficiency, requirement for small amount of equipment) are requirements for a good process test, none are sufficient if item validity is not demonstrated. That is to say, an item testing the process of controlling variables is valid when a person who can actually control variables gives a correct response and when one who can't gives an incorrect response. In test construction, validity is frequently appraised in terms of content validity.

Validity

Content validity is related to how well the content of the test samples the subject matter about which inferences are to be made. The method of determining this relationship is to obtain the opinions of persons knowledgable in the content area sampled. It has generally been left up to these persons to examine each test item and decide whether the items will elicit the appropriate response from the learner. The relationship for each item is then determined by looking at the opinions from each person.

If the test items are to measure specific facts, or the recall of specific information, content validity is probably sufficient. If, however, the items are to measure something other than factual retention, such as whether the learner is able to control variables or interpret data, content validity is, in the author's opinion, not sufficient. There is a need for test items that correlate with a student's abilit

of tes

validi

test i

extern

studen

measur

defini

and th

observ

behavi

test i

Recent

test o

scienc

1968)

used w

ted in

the te

Even t

tent v

validi

ability to actually perform the processes in question.

The validity that is most desirable for this type of test is criterion-related validity. Criterion-related validity pertains to a relationship between the scores on test items and the behavior exhibited on an independent external measure.

A test over processes of science should use the student behavior from a real situation as their external measure. The actual student behavior can be determined by defining the processes being evaluated in behavioral terms and then constructing situations in which the evaluator can observe the student's behavior. The individual student behavior could then be used as the criterion for validating test items on that process.

Recent Attempts in Evaluating Processes

There have been two major attempts to develop a test over scientific processes since AAAS published their science program Science--A Process Approach.

The Test of Science Processes, (Tannenbaum, R.S., 1968) is an instrument over eight processes which are to be used with junior high school students. The validity reported in this study is content validity obtained by submitting the test items to a jury of science education experts. Even though this report shows considerable evidence of content validity, there is no evidence of criterion-related validity. While Tannenbaum realized the importance of criteri

to give

to demo

subject

cesses

Jean, 1

suring

student

a group

criteri

provide

stated

proach

"Compet

the att

was des

individ

teacher

validity

to solv

is desig

involved

instrum

The "Ap

criterion-related validity, he made only a limited attempt to give some evidence of it. His investigation, undertaken to demonstrate criterion-related vality, used the teacher's subjective rating of the student's ability to use the processes of science as the external criterion measure.

The other test, <u>Basic Science Process Test</u>, (Beard, Jean, 1970) was designed to measure the processes of measuring and classifying with first, second, and third grade students. Even though the test meets the requirements of a group test, it, too, uses content validity instead of criterion-related validity.

The science program <u>Science--A Process Approach</u> provides two means of evaluating students' attainment of the stated objectives. Each lesson of <u>Science--A Process Approach</u> contains an "Appraisal" (group assessment) and a "Competency Measure" (individual assessment).

The "Appraisal" is designed to evaluate, as a group, the attainment of the specific skills for which the lesson was designed. It requires the student to work either individually or in small groups on a task set up by the teacher. Even though there is evidence of criterion-related validity (observing the student either solving or helping to solve the task which involves the process) the instrument is designed for students familiar with the vocabulary or involved in <u>Science--A Process Approach</u>. This makes the instrument usable for only a limited population of students. The "Appraisal" must also be administered by someone who is familiar with the behaviors that define the processes being tested.

The "Individual Competency Measure" is designed to evaluate science processes in a one-to-one situation. The student is presented with a series of tasks set up by the evaluator which would require the student to use the process in question. The validity of the competency measure is determined by the correspondence between the behavioral objectives and the tasks required by the competency measure. The competency measures were developed to give observational evidence of a student's attainment of the science process in question.

The "Individual Competency Measure" indicates the student can perform the process when they can successfully complete all the tasks of the competency measure. For example: A student can control variables if he successfully handles the set of tasks identified in the "Individual Competency Measure" for the process of controlling variables.

One of the problems with the "Individual Competency Measures" is that they were designed to be used with students who are familiar with the vocabulary of <u>Science--A Process Approach</u>. The competency measure also requires the administrator to be familiar with the required behaviors which define the processes. Another problem is that they were not designed for group administration.

Purpose

The purpose of this study was to develop a pool of objective test items for use with seventh grade students to measure the ability to perform two of the integrated processes. The two processes selected were Controlling Variables and Interpreting Data. These processes have been defined with behavioral objectives and measured by the use of "Competency Measures" in Science--A Process Approach. The criterion measures for this study were selected from those behavioral objectives and "Individual Competency Measures". By comparing the analysis data for each objective test item to the data from the respective "Individual Competency Measures", the items which adequately test the processes were determined. This study was so designed that the items which were accepted did, in fact, require the students to be able to perform the process. The criterion measure was given and scored, the students in the upper and lower 27% were determined, and from these groups the index of item discrimination[7] was calculated. Only those items which produced an index of item discrimination of .20 or greater were accepted. This level, .20, was chosen because scores on an external criteria was used in computing discrimination. This is explained in greater detail in Chapter III.

In particular, the objectives of this study are to develop objective test items over the processes of controlling

---

[7]The ability of a test item to differentiate between persons possessing much of some trait and those possessing little.

variables and interpreting data that:

1. will test the same behaviors as the selected "Individual Competency Measures" from Science--A Process Approach.
2. will be capable of administration by anyone interested in evaluation of students' acquisition of the specified processes.
3. can be used in the development of an instrument for use in a group testing situation.
4. can be administered with a minimum of manipulative material.
5. uses a vocabulary suitable for use with seventh grade students even if unfamiliar with the vocabulary in Science--A Process Approach.

Design

To fulfill the purposes of this study a combination of two designs was chosen. The two designs were: One-Shot Case Study and the One-Group Pretest-Posttest. The combination of these two designs implies the technique of making two separate observations after the group of students have been exposed to some experimental treatment. In this study, the prior exposure to the science program Science--A Process Approach will act as the experimental treatment while the two observations will be the administration of the selected "Individual Competency Measures" and the group test "A Group Test of Four Processes."

Assumptions

In conducting this study, the following assumptions were made:

1. Paper and pencil items can be constructed to adequately measure the selected processes of science.
2. The "Individual Competency Measures" of Science--A Process Approach are valid measures of a student's ability to control variables and interpret data.
3. The students who have studied Science--A Process Approach in the school year prior to the one in which this study was conducted will have sufficient experience with the vocabulary and manipulative skills used in the "Individual Competency Measures" to be suitable subjects.
4. The choice of the index of item discrimination as the basis of evaluating the effectiveness of the test items is appropriate for this study.
5. No appreciable amount of learning will take place as a result of the administration of the "Individual Competency Measures" prior to the student response to the written test items.

Limitations

In conducting this study, the following limitations are made:

1. No attempt has been made to construct a test. Only a pool of test items that have a high index of item discrimination and that exhibit criterion-related validity have been constructed.
2. This study was limited to one group of 75 seventh grade students in one school in Okemos, Michigan. The students had received instruction in Science--A Process Approach the year before and was limited to the material covered during that year.
3. The "Individual Competency Measures" used and the test items constructed were limited to sub-tasks in which the sample could reasonably be expected to have competency.
4. This study was also limited to an upper middle class, suburban, professional bedroom community.

Overview of the Dissertation

In this chapter the background, problem, need, purpose, assumptions and limitations of this study were presented. Chapter II contains literature related to the measurement of educational processes and some of the current attempts to evaluate the processes of science.

Chapter III contains a detailed description of the procedures used in carrying out this study, including the selection of the sample, administration of the "Individual Competency Measures", development of the test items and the procedures for the analysis of data.

Chapter IV includes the results and interpretations of the analysis of data. Chapter V presents the summary, conclusions and implications for further research derived from this study.

CHAPTER II

REVIEW OF THE LITERATURE

Recent science curricula developments illustrate a shift in the recommended emphasis of science instruction from a body of scientific facts to helping the children to discover, acquire and apply the information gained. Much has been said concerning the merits of emphasizing the means by which these facts and concepts are obtained. Sund and Trowbridge point out that science of the future will need to be "dynamic and changeable in order to meet the demands of a rapidly accelerating scientific age."[8] The new science curricula are giving our students the "opportunities for increased laboratory work and application of inquiry methods for learning. They are directed to better understandings of how scientists work and how knowledge is obtained. More attention is given to the processes of science."[9] Sund and Trowbridge also point out that the new curriculum projects

---

[8] R. Sund and L. Trowbridge, Teaching Science by Inquiry in the Secondary School, Ohio: Charles E. Merrill Books, Inc., 1967.

[9] Ibid.

will help students "gain skills and knowledge needed to handle the problems of this age successfully."[10] Hurd and Gallagher stress that teaching the processes of science and its modes of inquiry "is providing young people with tools that are not only applicable in the content of science but to life in general. What is more, if children learn to use these processes of science, they will come closer to understanding the spirit of inquiry that characterizes science."[11] Thus, the development of student abilities necessary to carry on the processes of scientific inquiries seems to be a major goal of science instruction.

Cole[12] states that process education recognizes as its foremost objective, the processes which the learner needs if he is to acquire, organize and utilize in a productive manner the information that he gains.

In this chapter the author will present some of the literature concerning the measurement of educational processes, and particularly the processes of controlling variables and interpreting data. In addition, some of the current attempts to evaluate the processes of science will be presented.

---

[10]*Ibid*.

[11]P. Hurd and J. Gallagher, *op. cit*.

[12]Henery P. Cole, "Process Curricula and Creativity Development," *Journal of Creative Behavior*, Vol. 3, No. 4, Fall 1969, pp. ~~243-44~~.

What Are Educational Processes?

The educational processes that the learner needs carry various titles. Bruner calls them intellectual habits.[13] Piaget refers to them as the logical operations.[14] Skinner calls them self management behaviors,[15] while Gagne has referred to them as learned capabilities and intellectual skills.[16] Burns and Brooks[17] say that processes are specific mental skills which are any of a set of actions, changes, treatments, or transformations of cognitive or affective entities used in a strategy in a special order to achieve the solution of a problem associated with the learning act, the use of learned products, or the communication of things learned.

Gagne states that educational processes are not entities of verbalization. They are not content processes such as the processes of economics. Instead, they are the strategies, the behavioral capabilities, the psychological

---

[13]Jerome S. Bruner, The Process of Education, New York: Random House, Inc., 1960.

[14]Jean Piaget, Six Psychological Studies, New York: Random House, Inc., 1967.

[15]B.F. Skinner, The Technology of Teaching, New York: A. Century-Crofts, 1968.

[16]Robert M. Gagne, "Contributions to Human Development," Psychological Review, 75, (3), 1968.

[17]Richard W. Burns and Gary D. Brooks, "What are Educational Processes?" The Science Teacher, Vol. 37, No. 2, Feb. 1970, pp. 27-28.

processes used by the individual to deal with the content.[18] Gagne's recent writings offer one of the best definitions of processes. "Processes are not something mysterious which the learner acquires, but behavioral tendencies or habits in particular types of situations."[19] Processes are merely names applied to collections of behavioral capabilities which are observed or inferred in the student.

Why Teach Processes?

Several justifications for teaching processes have been stated by Bloom.[20] First, the world is changing so fast that it is impossible to predict what knowledge and information will be needed. Therefore, it is important to help the student acquire generalized intellectual abilities and skills which will serve him in many new situations. Second, the amount of knowledge is so great that it is impossible to instruct the students in anything but a small portion of what is known. Thus, one of the goals of education must be to help the students to acquire some of the more relevant and central information combined with those processes which will help him adapt and expand this limited

---

[18] Robert M. Gagne, "Learning Hierarchies," Presidential Address, Division 15, American Psychological Association, San Francisco, August, 1968.

[19] Gagne, "Contributions to Human Development," op. cit.

[20] Benjamin S. Bloom, Taxonomy of Educational Objectives, Handbook 1: Cognitive Domain, New York: David McKay Company, Inc., 1956, pp. 32-43.

knowledge that he acquires. Third, the acquisition of essential processes will contribute to an individual's ability to successfully solve problems. Fourth, processes are more transferable than specific knowledge. Bloom says that processes are more widely applicable than knowledge and if we are concerned with the problems of transfer, we would then select processes over specific knowledge.[21] Fifth, intellectual skills are retained longer than other types of learning. Bloom also states that "it would seem reasonable to expect greater permanence of learning for those outcomes of education which can be generalized and applied in a number of different situations throughout the individual's formal educational experience than those outcomes which are so specific that they are likely to be encountered only once or at the most a few times throughout the educational program."[22] There has been some evidence to show that specific information which is learned is subject to rapid extinction, but the intellectual skills are more permanent.[23,24,25]

---

[21] Ibid., pp. 38-42.

[22] Ibid., p. 42.

[23] H. Freud and N.D. Cheronis, "Retention in the Physical Science Survey Course," Journal of Chemical Education, 17, 1940, pp. 289-93.

[24] P. Frutchey, "Retention in High School Chemistry," Journal of Higher Learning, 8, 1937, pp. 217-218.

[25] Ralph W. Tyler, "Permanence of Learning," Constructing Achievement Tests, Columbus: Ohio State University, 1934.

Gagne uses all of these arguments as justification for the development of processes and adds at least one more for consideration. He points out that information can and is usually recorded somewhere. This means that, if needed, information can be readily accessible when needed. On the other hand, process skills are not so easily acquired since they must be developed.

It is this justification that has helped influence the shift in emphasis of science instructional materials in today's schools.

Identifying Processes

The assumption that there are basic processes of science is advanced by Gagne in The Psychological Bases of Science--A Process Approach.

1. The scientists' behavior in pursuing science constitute a highly complex set of intellectual activities which are, however, analyzable into simpler activities.
2. These intellectual activities (processes) are, as most scientists would agree, highly generalizable across scientific disciplines.
3. These intellectual activities of scientists may be learned, and it is reasonable to begin with the simplest ones and build the more complex activities out of them, since this seems to be in fact the way they are organized.

4. Accordingly, one can construct a reasonable sequence of instruction which aims to have children acquire process skills, beginning with the simplest kinds of observations, and building progressively through classifying, measuring, communicating, quantifying, organizing through space and time, to the making of inferences and predictions. As further building occurs, one finds it possible for the students to learn how to make operational definitions, how to formulate testable hypotheses, how to carry out experiments, and how to interpret data from experiments.[26]

The Commission on Science Education of the American Association for the Advancement of Science has identified eight basic processes and five integrated processes for their elementary school science program, Science--A Process Approach. The eight basic processes are: observation, classification, recognition and use of space/time relations, recognition and use of number relations, measurement, communication, inference, and predictions. The five integrated processes are: formulating hypothesis, defining operationally,

---

[26]American Association for the Advancement of Science, Commission on Science Education, The Psychological Bases of Science--A Process Approach, Washington D.C.; AAAS, 1965, (AAAS Miscellaneous Publication 65-8.)

controlling variables, interpreting data, and experimenting.[27]

For each of the processes, the Commission has developed many exercises. Each exercise contains a set of behavioral objectives that describe what the child will be able to do at the end of the exercise, a rationale, a vocabulary list, a list of needed materials, suggestions of how to originate the problem, a suggested instructional procedure, suggestions of how to help generalize experiences, and two evaluative methods called "Appraisals" and "Competency Measures".

Two of the integrated processes, controlling variables and interpreting data, as defined by the Commission on Science Education of the American Association for the Advancement of Science, have been chosen for this study.

Defining the Two Processes

Controlling Variables

Variables are the conditions, properties, or factors that may influence the operation of something in a system. The variables are identified by observing the components and properties that are included in the system. To determine whether a variable influences the operation of part of a system, all other variables must be held constant while the one variable being tested is manipulated. Therefore, if a student can control variables, he will be able to:

---

[27]American Association for the Advancement of Science, Science--A Process Approach, New York: XEROX Education Division, 1967.

1. IDENTIFY variables which may influence the behavior or the properties of a physical or biological system.
2. IDENTIFY variables which are held constant or manipulated in an investigation or an experiment.
3. DISTINGUISH between conditions which hold a given variable constant and conditions which do not hold a variable constant.
4. CONSTRUCT a test to determine the effects of one or more variables on a responding variable.
5. IDENTIFY and NAME variables which were not held constant in the description of an investigation, although they varied in the same way in all treatments or were randomized.[28]

Interpreting Data

The process of interpreting data requires the integration of the skills learned in the basic science processes plus ingenuity on the part of the interpreter. The ability to use this process is desirable for the child as a future citizen, since much of his daily activity will be concerned with interpreting data. If a student can interpret data, he will be able to:

1. DESCRIBE in a few sentences the information

---

[28]American Association for the Advancement of Science, Science--A Process Approach, Commentary for Teachers, Commission on Science Education, American Association for the Advancement of Science, 3rd Experimental Edition, 1968.

shown in a table of data or graph.

2. CONSTRUCT one or more inferences or hypotheses from the information given in a table of data or graph.
3. CONSTRUCT one or more inferences or hypotheses from a comparison of the information in two or more related tables of data or graphs.
4. DESCRIBE certain kinds of data, using the mean, median, range, and frequency distribution; and CONSTRUCT predictions, inferences or hypotheses from this information.
5. CONSTRUCT inferences or hypotheses from pictorial data.
6. DISTINGUISH between linear and nonlinear relations, APPLY A RULE to find the slope of graphs and linear relations, and DESCRIBE the information provided by the slope.[29]

Recent Efforts in Evaluating Science Processes

The problems of developing methods of evaluating the newer curriculum projects have not been given the attention they deserve. Evaluation should be an integral part of curriculum development. Primarily it is important that the instruction be evaluated by means of instruments which measure the competencies that are included within the goals and philosophy of the program. However, evaluation is more

---

[29] *Ibid.*

than testing and examining. It is studying and assessing the total growth of the students in relation to their potential and the objectives of the program.

In a classroom that is content-oriented, the major goal is the student's acquisition of scientific facts and principles. Evaluation of this goal is possible through tests such as true-false, multiple choice, short answer and essay questions that test facts but do not necessarily require the processes used in determining those facts.

In a curriculum that includes science processes as part of their objectives, the evaluation should reflect not only students' acquisition of facts and principles, but also their ability to use the processes of science.

Recently there have been several attempts to assess the student's ability to use the processes of science. They are:

1. the "Competency Measures" of the American Association for the Advancement of Science's program Science--A Process Approach.[30]
2. the "Appraisals" of the American Association for the Advancement of Science's program Science--A Process Approach.[31]
3. Processes of Science Test by the Biological

---

[30]AAAS, 1967, op. cit.

[31]Ibid.

Science Curriculum Study group.[32]

4. Basic Science Process Test by Jean Beard.[33]

5. Test of Science Processes by Robert Tannenbaum.[34]

6. Stanford Achievement Test, 1964 edition.[35]

7. Metropolitan Achievement Test, Advanced Science Test.[36]

8. Sequential Test of Educational Progress: Science (STEP Science).[37]

There has also been an attempt to assess the science processes of teachers. Two of the major works in this area are:

1. Science Process Test for Elementary School Teachers by Evan Sweetzer.[38]

2. Science Process Measure for Teachers by the

---

[32]Biological Science Curriculum Study, Processes of Science Test, New York: The Psychological Corporation, 1962.

[33]Jean Beard, Group Achievement Tests Developed for Two Basic Processes of AAAS Science--A Process Approach, Ed.D. Thesis, Corvallis, Oregon State University, 1970.

[34]Robert Tannenbaum, The Development of the Test of Science Processes, Ed.D Thesis, Columbia University, 1968.

[35]Oscar K. Buros, (ed.), The Sixth Mental Measurements Yearbook, New Jersey: The Gryphon Press, 1965.

[36]Ibid.

[37]Ibid.

[38]Evan Sweetzer, Science Process Test for Elementary School Teachers, Michigan State University.

American Association for the Advancement of Science.[39]

Competency Measure

The Commission on Science Education of the American Association for the Advancement of Science has developed a process oriented elementary school science program through more than nine years of extensive research, experimentation, testing, try out, revision, and evaluation on a nationwide scale. This program, Science--A Process Approach, undertakes development of the scientific processes. For each of these processes, the commission has devised a list of behavioral objectives that describe exactly what the learner will be able to do. One of the evaluative devices of this program, the "Competency Measure", will be reviewed now while the other device, the "Appraisal", will be reviewed next.

> The competency measure consists of tasks intended to assess the achievement of the objectives for each exercise. The effectiveness of the measure is directly related to the accuracy with which the behavioral objectives of each exercise are translated into performance tasks that sample the behaviors described as objectives for that exercise. The validity of the competency measures rest upon the demonstration of the existence of the one-to-one correspondence between the behavioral objectives of the specific process exercise and the tasks required by the competency measure for that exercise.[40]

---

[39]American Association for the Advancement of Science, 1967, op. cit.

[40]American Association for the Advancement of Science, An Evaluation Model and Its Application: Second Report, Commission on Science Education, American Association for the Advancement of Science, 1968 (AAAS Misc. Publication 68-4).

This instrument is administered individually and is constructed around the stated behavioral objectives for the exercise. In this type of test, the teacher, or some other competent person, using the behavioral objectives as a standard, actually observes the student's behavior to determine whether the student can perform the process being tested.

The competency measure from Exercise 2 under the heading "Classification" will be used as an example. This exercise deals with constructing and demonstrating the use of different classification systems. It includes the classification of some objects on the basis of whether or not the objects can be used in a given way and then constructing a second classification of the same objects, called reclassifying, on the basis of another property. The behavioral objectives for this exercise are clearly stated in terms which prescribe the sorts of behaviors that the students must exhibit in order to demonstrate competency in this phase of classifying. The "Individual Competency Measure" which is used to measure the student's achievement of these objectives involve a group of objects (a ball, a block, a slice of bread, a grape, an orange, a ladle, and an empty tin can). These objects are placed in front of the child and his first task is to organize them into two groups depending on whether or not they can be eaten. After this the child is told to reclassify the objects on the basis of whether or not he could use the objects to carry water. Finally, the teacher makes an inclined plane and asks the child to reclassify the objects once more on the basis of whether or not they can roll down the inclined plane.

This example is taken from Part A of Science--A Process Approach and is usually presented in kindergarten.[41]

Even though the "Individual Competency Measure" is a valid test of the objectives of the exercise, there are some disadvantages to this method of evaluation. First, the competency measure must be administered by someone who is familiar enough with the processes to be able to interpret whether the student has performed the behaviors specified by the behavioral objectives. Second, it is almost impossible for a teacher to spend that much time in evaluating her class because this method of evaluation demands a low student-teacher ratio. Third, a considerable amount of equipment may be required because the student is frequently required to set up experiments and then carry them out. The evaluator must have on hand any equipment for which the student may ask.

Appraisal

The "Appraisal" is designed to evaluate, as a total class, the attainment of the specific skills for which the lesson was designed. They are set up so students work either individually, in small groups or as an entire class on a set

---

[41]From Science--A Process Approach, Part A, New York: XEROX Education Division, 1967.

of tasks related to objectives of the exercise. Exercise 12 under the heading "Observation" will be used as an example. This exercise, Observing Mold Gardens, is designed so each child should be able to do the following:

1. state the color, shape, and size of a few kinds of molds.
2. construct a chart of the class record of the color, shape, and size of one mold colony at intervals during its cycle.
3. state the color, shape, and size of the mold colony at intervals during its cycle by reading a chart.[42]

The "Appraisal" for this exercise says,

> Show the class a vegetable or fruit which has a mold colony on it. Ask children to describe the growth. After several children have responded and the teacher has a sampling of responses, each child might be asked to draw the object and include a careful representation of the mold colony.[43]

Even though the appraisals are designed around the specific objectives of the exercise, there are some disadvantages to this method of evaluation. First, the "Appraisals" must be administered by someone who is familiar enough with the processes to be able to interpret whether the student has performed the behaviors specified by the behavioral objectives. Second, with individual members of the class responding to the question, it is almost impossible to determine the extent to which each child has mastered the objectives of the lesson.

---

[42]From Science--A Process Approach, Part B, New York: XEROX Education Division, 1967.

[43]Ibid.

Process of Science Test

The Biological Sciences Curriculum Study group in developing this test was concerned with the students' knowledge of, and ability to use, the methodology of science. A list of their abilities include judging facts, interpreting data and evaluating experimental designs. The test consists of forty multiple choice items. All of the questions were chosen from the biological sciences, but the authors indicate that this is incidental to the main purpose of the test. They feel that any knowledge of biology is unnecessary for a high score on the test. Therefore, they state that the test may be useful in courses in biology and also in other science courses that stress the processes of science. The items are multiple choice and suitable for group administration. The test seems to be aimed at the tenth grade student.

The real disadvantage of this test is the absence of criterion-related validity. It has not been validated against the actual behaviors that would indicate the student's ability to use the processes that are being tested.

Basic Science Process Test

Jean Beard developed a test over two of the basic processes as defined by Science--A Process Approach. This test is aimed at the early elementary school child. It is administered in a series of 35 mm colored slides which illustrates laboratory situations involving basic science processes. The slides are in sequence with a synchronized tape recording which provides oral instruction for the children, states the problem to be considered, and changes the slides. Each child has his own answer sheet which he marks as directed by the recording. This procedure was used as an attempt to assess science process skills of primary pupils who had not yet perfected their reading and writing skills.

Even though the test is appropriate for group administration and can be administered by anyone, there is still the question of validity. The validity of this test was determined by a jury who evaluated each item and made a decision concerning the probability of the abstract item eliciting the appropriate information from the children --a similar procedure to that used by authors of standardized tests. The items were not evaluated against the actual behaviors of the children.

Test of Science Processes

Tannenbaum has developed a paper and pencil group test on some of the processes of science. The processes he selected were: observing, comparing, classifying, quantifying, measuring, experimenting, inferring, and predicting. For each process, a list of behaviors that the student must exhibit in order to demonstrate competency were specified in detail. These statements were submitted to experts for their opinions and validation before final writing. The instrument was then administered to 3,673 students who were selected to include all ability levels and a wide range of socioeconomic backgrounds. The results of this administration were used to create norms for total score and for each of the eight subscores. The norms are reported by grade, sex, and by urban versus suburban, and for the total sample within each grade. This test reports a high total reliability (o.91) and reliability-coefficients for each of the eight subgroups ranging from 0.30 to about 0.80. It is important to realize that these reliabilities on the subparts of the test (using Kuder-Richardson 20 formula) are very low and must be used with caution.

The study reports considerable evidence of content and curricular validity of the test. While there is evidence of validity, submitting test items to experts for validation does not guarantee that the items are measuring the student's ability to use the process. Tannenbaum says that the criterion-related validity of this test is very difficult to assess because "this is the first attempt to measure this age level student's ability to use science processes."[44] The criterion-related validity could be attempted if the standard were the student's ability to perform the science processes in an actual situation where the student is faced with a science phenomenon. If, in this actual situation, the student's behaviors define the

---

[44]Tannenbaum, op. cit.

processes being tested, validity could easily be established. Tannenbaum reports a small investigation undertaken to demonstrate the criterion-related validity where the criterion was the teacher's rating of the student's ability to use the processes of science. "The results of this investigation were not unequivocal. However, they do give some indication of a degree of criterion-related validity. More conclusive evidence of this must await much more extensive investigation."[45] A major question that needs to be answered in order to determine how much confidence can be placed in the "indication of criterion-related validity" stems around the background of the teachers in determining the processes of science. In this case, the validity still relies on someone's interpretation of the processes of science.

Stanford Achievement Test

M. Byran points out that the authors of the *Stanford Achievement Test* have sought to insure content validity by "Examining appropriate courses of study and textbooks as a basis for determining the skills, knowledge, understandings, etc. to be measured."[46] The author could find no indication that the test items were based on a performance criterion.

Metropolitan Achievement Test

W. Cooley states that the authors of this test "have

---

[45] *Ibid*.

[46] M.M. Byran, in Buros, *op. cit*.

been far too concerned with the standardization process, while item content and validity appear to have been almost completely ignored."[47] He goes on to say that the tests are "open to very serious criticism regarding the nature of the items and the knowledge and abilities assessed."[48]

Sequential Tests of Educational Progress: Science (STEP Science)

The STEP Science test is designed to measure the student's knowledge of basic scientific concepts and his problem solving skills. The test items were written by a committee who worked together for several weeks so they could "criticize each other's work, maintain continuity from level to level and 'rub their ideas together'."[49] Even though no coefficient of validity is reported, the authors indicate they were concerned about content validity.

All three of these standardized tests have relied upon a group of interested people to determine content validity. It has been up to these people to determine whether the test items measure the skills and concepts used in existing science curricula.

Even though a group of interested people have stated that these test items will elicit the appropriate responses, there is no guarantee. To determine whether these test items

[47]W.W. Cooley, in Buros, Ibid., p. 877.

[48]Ibid.

[49]Ibid.

do in fact measure the skills and processes they were designed to measure, they must be compared to the actual ability of the student to perform the skills and processes.

<u>The Other Instruments</u>

There are three other works in the field of science processes. <u>Science Process Test for Elementary School Teachers</u> and <u>Science Process Measure for Teachers</u> are both designed to assess teacher's knowledge of and ability to use processes. The other instrument, <u>Wisconsin Inventory of Science Processes</u>, seems to be designed to measure knowledge of, and not necessarily ability to use, the processes of science.

Problems with Evaluating Processes

When one considers the evaluation of processes, one must be cautious of expecting a ready made test which is appropriate for group administration. There are several problems associated with assessing the processes of science.

Ebel gives two reasons why items written to measure the higher mental processes have failed to assess the processes. "One is that they may be quite difficult and thus call for more than ordinary examinees are capable of delivering....Another is that they may involve fairly complex situations, which require many words to describe and may present the examinee with problems of comprehension and interpretation which may be irrelevant to the main purpose of the examination."[50] The study presented here has attempted to deal with these problems by providing both visual and oral information about complex situations and by using vocabulary which students should be expected to know.

Another problem associated with the evaluation of science processes is that of time. To determine whether a student can perform a particular process, he must be able to demonstrate the behaviors which define that process. To successfully evaluate science processes, the evaluator needs to work with each child one at a time. There is just not enough time for a teacher, or evaluator, to assess the competency of all children in all the different processes of science this way. If a process test is to be developed, it must be more efficient. The most efficient way would be to develop an instrument which could be administered to a group of students at one time.

Another problem with assessing the processes of science is the amount of equipment that is needed. Usually, evaluating the student's competency requires situations which permits one to observe the student working with equipment. As the student investigates the situation he must have the freedom to work with all and any equipment he feels that he needs to successfully complete the tasks. There must be some way of eliminating the amount of equipment which is needed. There needs to be a test which can

---

[50] Robert L. Ebel, Measuring Educational Achievement, New Jersey: Prentice Hall, Inc., 1965, p. 52.

be administered by anyone, does not take too much time, and requires little or no material.

One solution to the above problems is to produce a pencil and paper test that is objective and uses little or no equipment. As mentioned earlier, several process tests are available (Beard, Tannenbaum, Buros, etc.). Associated with the production of such a test, however, is the difficulties experienced in trying to produce test items that are, indeed, valid.

To overcome this, a decision was made to use an external criterion as a means of validating the test items. Horrocks and Schoonover in their book Measurement for Teachers[51] describe the requirements of a criterion. These requirements are: 1) true outcomes of the construct in question, 2) observable, 3) measurable in some quantitative fashion, 4) readily definable, and 5) agreed upon by the individuals concerned with establishing the behavior as a criterion. The "Individual Competency Measures" from Science--A Process Approach appear to meet these conditions as an adequate criterion and will thus be used as the external criterion measure.

Summary

The current trends in science education have placed an emphasis on the development of the processes of science.

---

[51]John E. Horrocks and Thelma I. Schoonover, Measurement for Teachers, Ohio: Charles E. Merrill Publishing Company, 1968, p. 70.

There is at least one program that is committed to the instruction of the processes of science. This program, Science--A Process Approach, has also developed a method to evaluate the student's acquisition of these processes. This method, called the "Competency Measure", requires administration in a one-to-one situation by someone who is familiar with the processes being tested. Since the evaluation requires a one-to-one situation, it is not really feasible for the teacher to evaluate her entire class.

There have been other attempts to evaluate the student's abilities to use the processes of science. All are suitable for group administration by anyone, and four are geared for junior high school students. The real problem is their failure to show a relationship between a score and some recognized standard of performance. Even though the items have been submitted to experts in elementary and junior high school science education, it is important to remember that this does not guarantee the relationship of a student's test score to his ability to use the actual processes. This study is an attempt to develop a method of evaluation possessing a direct relationship between students' scores on a group test and their scores on the standard, the "Individual Competency Measures" from Science--A Process Approach.

CHAPTER III

METHODS AND PROCEDURES

Overview

The purpose of this study was to create objectively scored, large-group administered, criterion validated test items that could form the basis of a test to measure student's performance on the processes of controlling variables and interpreting data.

The work in this study is divided into two parts. The first, the development of the test items, including writing, critiquing, and rewriting items. It also includes the development of a series of colored slides to support some of the test items. The second part of this study is the actual testing of the items. This section includes the selection of the population and sample, instrumentation, and the analysis procedure.

Validation Design for the Items in a Group Test of Four Processes

The design for the validation portion of the study involved selecting a sample that had previously experienced Parts 4 and 5 <u>Science--A Process Approach</u>.

This experience was considered necessary because the criterion measures which were administered to the sample contained vocabulary which could be unique to this science program.

Two observations were obtained on the sample: 1) the criterion measure, selected "Individual Competency Measures" from Science--A Process Approach and 2) the test items from the group measure, "A Group Test of Four Processes". The first observation period was necessary to establish which students could and which students could not perform the processes in question. The second observation period was administered immediately after the first observation period to keep the confounding variable of maturation to a minimum. Detail concerning the administration of the criterion measure and the group measure will be presented later in this chapter.

Besides the confounding variable of maturation, a question arises concerning the "test wise" effect, or the amount of learning which might occur by taking a test. It was felt that if learning did occur during the observation periods, it would more likely occur during the administration of the criterion measure. Since the criterion measure was a situation where students were presented with a problem and allowed to work on the solution until they were satisfied with their results, any learning that did occur while working on the problem would be reflected in their scores on the criterion measure as well as their scores on the group measure, the second observation period.

Based on the assumption, if the criterion measure were given last, learning would not be reflected equally in the two measures. Because of this, it was decided not to split the students into two groups and alternate the administration of the two instruments.

Development of Test Items

To develop items for "A Group Test of Four Processes" on the processes of controlling variables and interpreting data, the following procedure was used:

1. The author reviewed all the material in Science--A Process Approach, Parts 5, 6 and 7.
2. Abstract items were then developed around situations which might be within the scope of junior high students.
3. The series of items were submitted to a group of faculty and graduate students at Michigan State University's Science and Mathematics Teaching Center for the purpose of establishing face validity. They were asked to make comments and suggestions as to whether these items would elicit the appropriate responses. The members of this group were: Dr. Richard J. McLeod, Associate Professor of Science Education; Darrel W. Fyffe, Donald Maxwell, Kenneth Calkins, and Howard Hetzig,

graduate assistants in science education; and Dr. Glenn D. Berkheimer, Associate Professor of Science Education.

4. Those items which the group felt were inappropriate were deleted while the items for which the group made suggestions were revised. The retained items were then reviewed by Dr. Richard J. McLeod and this author to determine the most appropriate means of presenting them to the students. According to Ebel, items written to measure the higher mental processes may fail because they "...may involve fairly complex situations, which require many words to describe and may present the examinee with problems of comprehension and interpretation which may be irrelevant to the main purpose of the examination."[52] With this in mind, it was felt that 25 of the 43 items would be more successful if they were presented with the aid of some visual materials. Because of the nature of the situations to be developed, they seemed to lend themselves to be presented with the aid of colored slides and an oral script. The remaining 18 items were judged sufficiently simple to be presented in the

---

[52]Robert L. Ebel, op. cit., p. 52.

usual paper and pencil form without the aid of a visual presentation.

5. Slides were developed for those items where it was deemed appropriate. The series of slides, along with accompanying items were then reviewed by McLeod and the author. Suggestions and revisions were made. The author then revised all items and placed them (the items and slides) in a sequence.
6. The same procedure of multiple critique and revision was followed for the non-visual items.
7. All the proposed items, those with a slide sequence and those without, were alternated with the items produced by Fyffe.
8. The total series was then reviewed by McLeod, Berkheimer, Fyffe and the author for the final revisions.
9. The items were revised and "A Group Test of Four Processes" was assembled. It is most important to note that this represents only an assemblage of test items and not a finished test. At this time, however, the items were judged to possess content validity.

The first 42 items of the 79 item test were developed around several different visual situations and recorded on

35 mm color slides. Of these 42 items, 12 were on the process of controlling variables and 13 were on the process of interpreting data. A script was prepared and read orally as the students progressed through the items. Of the remaining 37 items (without color slides), six were on the process of controlling variables and 12 were on the process of interpreting data. These items were presented in the manner that is most commonly used for standardized tests, presenting the test booklet and letting the students work at their own rate through the items. The entire script, test, and prints of slides are presented in Appendix A. Since this was a power test and not a speed test, the students were allowed sufficient time to finish. All the students had finished in 75 minutes and most finished in 70 minutes. Each item was designed to require the use of one particular science process and was prepared in a style and language that would make it suitable for use in schools where Science--A Process Approach is not used.

Testing of Items

The testing of the items developed included not only the actual process of testing but the selection of the population and sample.

Population and Sample

The criterion measures use terminology that is unique to Science--A Process Approach. Therefore, it was important to select the sample from a population that had been exposed to that science program to avoid a communication bias in the materials. The population for this study consisted of junior high school students in grade seven who had received instruction in Science--A Process Approach in the sixth grade.

The sample used in this study was selected from students who had attended the Kinawa Middle School in Okemos, Michigan, during the sixth grade. These students received, as part of their sixth grade science program, instruction from the science program, Science--A Process Approach, Parts 4 and 5. The seventh grade students in this school are grouped into two clusters. The school principal stated that the students are assigned to one of the two clusters without regard to ability or background. The clusters are then divided into three separate classes. These classes, science, social studies, and English, share a three hour block of time which allows the three teachers to arrange the three hours in any manner they wish. Through an agreement reached with the principal and teachers, one cluster of students was assigned to us to use as our sample (N=59). A working agreement was reached with the science, social studies and English teachers to periodically take one or more students from their classes for testing purposes.

The school chosen for this study, the Kinawa Middle School, is in the upper middle-class community of Okemos, Michigan. Okemos is a suburban, bedroom community that is located three miles from the campus of Michigan State University. This community contains a large share of the University's faculty and married graduate students. The public school system is well equipped and financially well supported.

Instrumentation

Two instruments were administered to each student in the sample. They were: 1) selected "Individual Competency Measures" from Science--A Process Approach, and 2) "A Group Test of Four Processes". The first instrument, the selected "Individual Competency Measures", consisted of the following:

1. two of the "Individual Competency Measures" for the process of controlling variables: a) Controlling Variables 7 consisting of four tasks,[53] and b) Controlling Variables 9 consisting of eight tasks.[54] There were a total of twelve tasks for the process of Controlling Variables.
2. three of the "Individual Competency Measures" for the process of interpreting data: a) Interpreting Data 9 consisting of six

---

[53]American Association for the Advancement of Science, Science--A Process Approach, Part Six, (AAAS Miscellaneous Publication 67-11), 1967, pp. 97-98.

[54]Ibid., pp. 140-142.

tasks,[55] b) Interpreting Data 12 consisting of three tasks[56] and, c) Interpreting Data 13 consisting of six tasks.[57] For the process of interpreting data, fifteen tasks were presented to the students.

These five "Individual Competency Measures" were chosen because they: 1) were representative of the activities for the processes in question, 2) contained a variety of tasks which represents the skills the sample could reasonably be expected to possess based on their previous year's experience, 3) did not limit the student to only one situation where recall could play a big part in success, and 4) used materials and equipment with which the students were familiar and which were readily available.

These measures were administered individually in an interview type situation according to the guide lines described in Science---A Process Approach. One score for each process was recorded which represented the total number of tasks each individual completed correctly for each process.

The students in this sample were also tested individually on two other processes which were part of the

---

[55] Ibid., pp. 82-83.

[56] Ibid., pp. 227.

[57] Ibid., pp. 247-248.

research of Darrel W. Fyffe.[58] The average time for the total individual testing of the four processes was about 90 minutes per student with controlling variables and interpreting data accounting for an average of 55 minutes of the 90.

The second instrument used was "A Group Test of Four Processes". This test was developed in cooperation with Darrel W. Fyffe whose primary concern was in developing items to test the processes of formulating hypothesis and defining operationally. The test consisted of 79 items of which 18 were on the process of controlling variables, 25 were on the process of interpreting data and the remaining 36 were on the other two processes, formulating hypothesis and defining operationally. The seventy-nine items were either multiple choice or numerical fill-in. An example of a multiple choice item would be item 30 and an example of a numerical fill-in would be item 5. All items can be found in Appendix A.

The reader should note that the purpose of this study was not to develop a test but test items that correlate with the student's ability to perform the processes of controlling variables and interpreting data as measured by the "Individual Competency Measures" from Science--A Process Approach.

---

[58]Darrel W. Fyffe, The Development of Test Items for the Integrated Science Processes: Formulating Hypothesis and Defining Operationally, unpublished dissertation, Michigan State University, 1971.

Administration of Instruments

The Kinawa Middle School is equipped with a faculty hallway that connects the science rooms with storage area, conference rooms and the professional library. The facilities that were provided for our individual testing were: one of the conference rooms, and two small table and chair sets in the hallway next to a sink that had hot and cold running water. The equipment used was stored in a small office.

Because of the schedule of the cluster of students, all testing was conducted in the morning. The students were available before lunch during the second, third, and fourth class periods. Each class period is 50 minutes with a five minute change period. For the individual testing, each researcher was able to draw one student at a time from the science classes.

To facilitate efficiency, each of the researchers had familiarized themselves with all the "Competency Measures" prior to testing and enough equipment was on hand so both researchers could be testing the same process at the same time. Generally this was not necessary. Each researcher was able to test our own particular processes, but at times both administered the same tests. In a number of cases the scoring was checked with the researcher of concern to assure that all scoring was uniform.

The individual testing was begun on April 13, 1970, and concluded on May 29, 1970. Fyffe was able to test five mornings a week while the author tested four mornings a week. Most students were completely finished with the "Individual Competency Measures" of all four processes within two periods of testing. Several took ten to fifteen minutes longer while a small number required a third period.

The group test was administered on June 3 and 4, 1970. On the first day, all three classes reached item 30 and, due to lack of time, the materials were collected. The group testing was completed during the second day. During the group testing, Fyffe handled the directions, controlled the lighting and observed the class for signs to indicate movement to another slide. This author read the script and advanced the slides. The classroom teacher allowed the researchers complete control of the classes during each period.

Each student was provided with a pencil, an answer sheet, and a test booklet. The answer sheet was marked as illustrated in the directions. Because of absences, only 56 of the original 59 students who had been tested individually completed the group test.

The students, teachers and administration were not informed of the individual or group test results.

Analysis Procedure

The student responses were transferred from their answer sheets to IBM H94060 answer sheets. The IBM answer sheet consists of two sections, the top part has space for 92 five choice items while the other section has 40 ten choice items. The responses to the 79 items of the group test were recorded on the top part of the answer sheet. The student's score on the "Individual Competency Measures" for the four processes was coded on the bottom part of the answer sheet for analysis.

Using the computer at Michigan State University's Test Scoring Office, the 56 papers were analyzed to determine the item analysis data based on the criterion measure, the "Individual Competency Measure". This meant that a separate item analysis was reported for each item for all four processes.

Two indices were of importance in this study, the index of item difficulty and the index of item discrimination. The index of item difficulty is the percentage of the total group marking a wrong answer or omitting the item. For example, if 64 percent of the students marked the question correctly, then the index of item difficulty would be 100 minus 64 or 36. The index of item discrimination is the difference between the percentage of the upper group marking the right answer and the percentage of the lower group marking the right answer. For example, if for a certain item, 80 percent of the upper group marked the correct response while 34 percent of the lower group marked the correct response, the index of item discrimination would be .80 minus .34 or .46. The upper group was defined as the 27% who received the highest scores for each process on the "Individual Competency Measures" while the lower group was defined as the 27% who received the lowest scores on the "Individual Competency Measures". Thus, it is important to note that success in the criterion measure determined the upper and lower groups. The students in the upper and lower groups for the process of controlling variables were not necessarily the same students as those in the upper and lower groups for the process of interpreting data. These groups were determined independently for each process in question.

To determine which items from "A Group Test of Four Processes" would be acceptable, a minimum level for the index of item discrimination was established. Usually, the index of item discrimination that has a value of .40 or greater is considered desirable. This must be looked at in the light in which the upper and lower groups are usually determined. The usual practice in determining this index is to define the upper and lower groups by the total score on the test. In this study, the scores on an external measure, the "Individual Competency Measures", were used as the criteria for separating the two groups. Therefore, the choice of the upper and lower groups were not affected by the responses to any test item. It was therefore suggested by Dr. Robert L. Ebel,[59] an expert in the field of testing, that for this study, any item having an index of item discrimination of .20 or greater should be accepted.

The other analysis reported in this study gives the Pearson product moment correlation coefficients for different pairs of the four processes from the "Individual Competency Measures" and "A Group Test of Four Processes". These correlation indices are important as they give further evidence of the ability of pencil and paper tests to measure student's ability to actively perform the processes in question. The correlation coefficients of interest for this study are:

1. the score on the "Individual Competency Measures" and those group items with a discrimination index of .20 or greater on the process of controlling variables.
2. the score on the "Individual Competency Measures" and those group items with a discrimination index of 0.20 or greater on the process of interpreting data.

Items exhibiting correlations that are significant at the .05 level or better will be accepted as measuring sub-

---

[59]Through a personal interview with Robert L. Ebel.

stantially the same thing as the corresponding competency measures.

Summary

With the goal of this study being the development of objectiviely scored test items that measure the processes of controlling variables and interpreting data, it was important to compare these proposed items against the actual behaviors that define the processes in question. The procedure followed was to set up a criterion measure, in this study the "Individual Competency Measures" from Science--A Process Approach, and then construct group test items over the same processes.

The test items show evidence of content (face) validity as they were evaluated and revised by science education experts before they were administered. In addition to content validity, these items show evidence of criterion-related validity using the "Individual Competency Measures" (actual student behavior) as the criterion measure. The difference in content and criterion-related validity is in the standard used for validation. Content validity uses the opinions from a jury of science education experts as the standard while criterion-related validity uses the actual student behavior as the standard.

There were two separate analyses run. First, the indices of item difficulty and discrimination were deter-

mined using the scores on the criterion measure as the basis. The second analysis gave the Pearson product-moment correlation coefficient for different pairs of the four processes of the "Individual Competency Measures" and "A Group Test of Four Processes". Of particular interest to this study were the coefficients for the individual and group scores on each process as further evidence of the ability of pencil and paper tests to measure students' abilities to actively perform processes of science.

CHAPTER IV

ANALYSIS OF RESULTS

The design of this study provides a mean to compare a student's performance on certain selected processes to their performance on pencil and paper group test items. Indices of item discrimination, item difficulty and correlation coefficients for the paper and pencil test items were produced and are reported in the following manner:

1. Data from the "Individual Competency Measures" for the processes of controlling variables and interpreting data.
2. Data from "A Group Test of Four Processes" for the processes of controlling variables and interpreting data.
    a. Data on non-acceptable items
    b. Data on acceptable items
3. Data relative to the items that were presented visually and non-visually.
4. Correlation coefficients which were significant to this study.

Individual Competency Measures

The following tables and graphs give the reader data concerning the frequency of scores on the "Individual

Competency Measures" for the processes of controlling variables and interpreting data.

Table 1

Frequency Distribution from the "Individual Competency Measures" for the Process of Controlling Variables (Total possible of 12)

Raw Score	Frequency	Cumulative Frequency	Percentile Rank
12	4	56	96
11	7	52	87
10	6	45	75
9	5	39	65
8	8	34	54
7	5	26	42
6	7	21	31
5	5	14	21
4	2	9	14
3	3	7	10
1	3	4	4
0	1	1	1

Range 0-12
Mean 7.4
Standard Deviation 3.1

Table 2

Frequency Distribution from the "Individual Competency Measures" for the Process of Interpreting Data (Total possible of 14)

Raw Score	Frequency	Cumulative Frequency	Percentile Rank
14	5	56	96
13	2	51	89
12	4	49	84
11	6	45	75
10	12	39	59
9	5	27	44
8	5	22	35
7	4	17	27
6	3	13	21
5	3	10	15
4	2	7	11
3	3	5	6
1	2	2	2

Range 1-14
Mean 8.8
Standard Deviation 3.3

Tables 1 and 2 show that the sample of students have a wide range of abilities to use the processes of controlling variables and interpreting data. The data shown in these distributions are both slightly positively skewed. This type of distribution does in fact enhance this study. If the scores on the "Individual Competency Measures" were negatively skewed, a high correlation on the group test items would be meaningless. It would only tell us that students who cannot perform the process also cannot correctly answer the items from the group test on that process. The following graphs, Figures 1 and 2, may help to visualize the distributions.

Frequency Distribution from the "Individual Competency Measures" for the Process of Controlling Variables (A Score of 12 was possible)

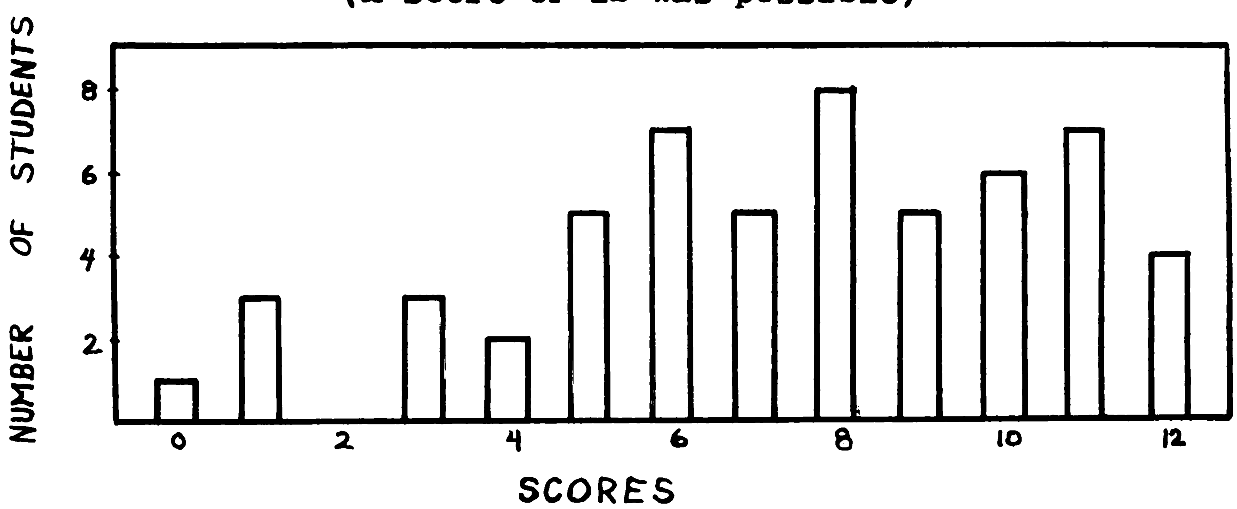


Figure 1

Frequency Distribution from the "Individual Competency Measures" for the Process of Interpreting Data (A Score of 14 was possible)

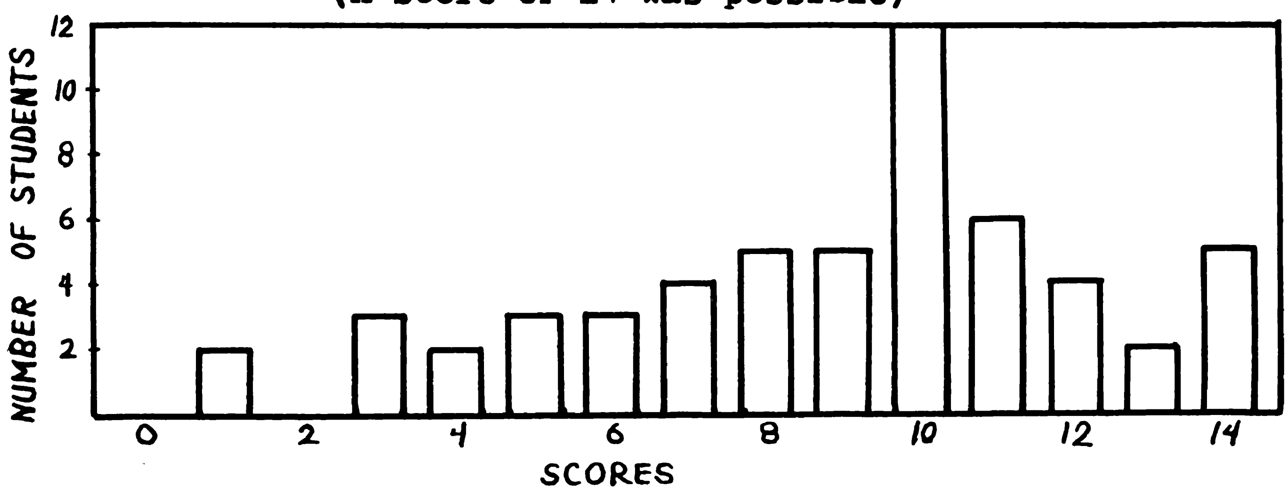


Figure 2

Even after observing that there is a real variance in the student's ability to control variables and interpret data, it could be assumed that not all the tasks from the "Individual Competency Measures" were contributing. Tables 3 and 4 show the percentages of students who responded correctly to each task of the "Individual Competency Measures".

Table 3

Percentage of Students Who Performed Each Task Correctly
Controlling Variables

Competency Measure	Task Number	Percent Correct	Index of Task Difficulty
Controlling Variables 7	1	63	37
	2	50	50
	3	59	41
	4	59	41
Controlling Variables 9	1	63	37
	2	80	20
	3	71	29
	4	66	34
	5	38	62
	6	54	46
	7	87	13
	8	54	46

Table 4

Percentage of Students Who Performed Each Task Correctly Interpreting Data			
Competency Measure	Task Number	Percent Correct	Index of Task Difficulty
Interpreting Data 9	1	48	52
	2	71	29
	3	43	57
	4	57	43
	5	86	14
	6	71	29
Interpreting Data 12	2	34	66
	3	82	18
Interpreting Data 13	1	87	13
	2	91	9
	3	89	11
	4	43	57
	5	41	59
	6	38	62

Tables 3 and 4 show that all tasks were contributing with only three falling below 40%. The range of task difficulty for all the tasks of controlling variables was 13 to 62 while the range of task difficulty for the tasks of interpreting data was 9 to 66.

The data presented concerning the "Individual Competency Measures" shows that all the individual tasks were contributing. Some tasks were easier than others but none were trivial. There were no tasks that all students got right and no tasks that all students got wrong.

The data also shows that the distributions for the processes of controlling variables and interpreting data are both slightly positively skewed. As stated earlier, this does in fact enhance the value of the criterion measure. If there were a large number of students who could not perform the processes in question, then subsequent analyses would be relatively fruitless.

Following the administration of the "Individual Competency Measures", the group measure ("A Group Test of Four Processes") was administered.

Group Measure

The following tables, Tables 5 and 6, show the percent of correct responses for each item and gives the corresponding indices of item difficulty and discrimination.

Table 5

Data from "A Group Test of Four Processes" for the Process of Controlling Variables

Item Number	Percent of Correct Responses	Index of Item Discrimination	Index of Item Difficulty
24*	79	.14	21
25	38	.40	63
27	71	.34	29
28*	23	.00	76
29	39	.40	60
30	46	.34	54

Table 5 (cont'd.)

31	32	.46	67
32	63	.47	38
33*	16	.14	84
34	41	.47	59
35	64	.40	35
36	55	.40	44
44*	41	.07 negative	59
50*	38	.13	62
63	57	.53	43
64	61	.66	39
65	54	.46	46
78	50	.33	50
MEANS	48.5	.33	51.5

* These items were rejected because of insufficient index of item discrimination.

Table 6

Data from "A Group Test of Four Processes" for the Process of Interpreting Data			
Item Number	Percent of Correct Responses	Index of Item Discrimination	Index of Item Difficulty
3	70	.26	30
4	29	.20	71
5	36	.20	64
6	57	.40	42
11*	61	.14	40
12*	23	.07	77
13	52	.60	48
23	32	.26	68

Table 6 (cont'd.)

26	20	.33	81
37	43	.53	57
38	70	.33	30
39	64	.20	36
40	39	.20	60
45	59	.33	40
46	64	.60	35
47	71	.46	29
48	36	.60	64
49	43	.20	57
66	52	.27	48
67	63	.26	37
68	64	.60	35
73	79	.53	21
74	84	.47	16
75	64	.40	36
77*	52	.13	48
MEANS	53	.34	47

* These items were rejected because of insufficient index of item discrimination.

Tables 5 and 6 show that while all items were meeting some degree of success, there were a few that did not achieve the predetermined index of item discrimination to warrant their retention. With the level of acceptance established at a discrimination index of .20 or greater, there were eight items that needed to be rejected. Five of these items were for the process of controlling variables and three items were for the process of interpreting data. These are indicated by an asterisk after the item number.

Non-acceptable Items

The item analyses data for the rejected items will be presented along with an attempt to analyze the reason for the item's failure. No attempt was made to retest the suggested revision because of the large number of acceptable items retained.

Item 24

Item 24 asks the student to analyze an investigation to determine which of the variables were changed. The original item read as follows:

24. Which of the following variables were changed in this investigation?
    1. amount of salt in the solution.
    2. temperature of the freezer.
    3. time required for the mixture to freeze.
    4. size of the container.

The response pattern for this item was:

Group	Foil Number 1	2	3	4	Omitted
Upper 27%	7%	87%	0%	7%	0%
Middle 46%	4%	77%	12%	0%	8%
Lower 27%	7%	73%	7%	7%	7%
Total	5%	79%	7%	4%	5%

Index of item discrimination .14
Index of item difficulty 21
The correct response is foil 2.

The foils for this item indicates that only one variable was changed during the investigation. Answering this question could then be a process of elimination. I would suggest that by changing the foils to include the possibility of more than one variable being changed, this item might be more acceptable. In doing this, I suggest revising the foils so the item reads:

24. Which of the following variables were changed in this investigation?

    1. temperature
    2. temperature and time
    3. temperature and amount of salt
    4. temperature, time and size of container

Item 28

Item 28 asks the students to look at a slide which shows three small beakers filled to the same level with a cloudy solution, one large bag of salt, and one large beaker with a cloudy solution. It is not stated that the three smaller beakers have been filled with the salt solution from the larger beaker. Assuming that the beakers did not have the same salt solution, there is no one correct answer since foils 1, 4, and 5 all become possibilities. The item response pattern shows the following:

		Foil Number					
Group		1	2	3	4	5	Omitted
Upper	27%	47%	0%	0%	27%	27%	0%
Middle	46%	31%	4%	12%	19%	27%	8%
Lower	27%	47%	0%	7%	27%	7%	13%
Total		39%	2%	7%	23%	21%	7%

Index of item discrimination .00
Index of item difficulty 76
The correct response is foil 4.

I feel it would be best to change the stem of the item so it includes a statement which indicates that the three smaller beakers were, in fact, filled with the same salt solution. The item might read:

28. By filling each of the three smaller beakers from the larger beaker, the students could investigate the effect of different:

    1. amounts of salt
    2. kinds of containers
    3. sizes of containers
    4. freezer temperature
    5. none of the above

<u>Item 50</u>

This item asks the students to identify the variable or variables which need to be held constant to test a certain hypothesis. The original item read:

50. Which of the following variables would you hold constant to test the hypothesis that the number of tablets dissolved in the water has no effect on the time needed to dissolve another tablet?

1. amount of water

2. number of saccharin tablets

3. temperature of the water

4. both amount and temperature of the water

The item response pattern is:

Group		Foil Number 1	2	3	4	Omitted
Upper	27%	20%	33%	13%	33%	0%
Middle	46%	8%	15%	27%	50%	0%
Lower	27%	20%	33%	7%	20%	20%
Total		14%	25%	18%	38%	5%

Index of item discrimination .13
Index of item difficulty 62
The correct response is foil 4.

A look at the response pattern indicates there is no one foil which the majority of students preferred. The pattern also shows a similarity in the way the upper and lower groups responded. All of this leads the author to feel that the item was confusing. It might help to change the stem of this item by including some extra words in an attempt to clarify the problem. The stem should specify that the tablets, which are mentioned, have already been dissolved in the water. Changing the stem in this way also suggests a change in the second foil to include the words "already dissolved". The item should read:

50. Which of the following variables would you hold constant to test the hypothesis: The number of tablets already dissolved in the water has no effect on the time to dissolve another tablet?

1. amount of water
2. number of tablets already dissolved
3. temperature of the water
4. both 1 and 3

There were five items for which this author was unable to suggest revisions which might make the items acceptable.

Item 33

Item 33 shows two rods: A, which is metal, two centimeters long and solid; and G which is plastic, eight centimeters long and solid. It then asks the students to determine which variable(s) you could test by rolling them down the same inclined plane. The item reads:

33. By rolling rods, A and G, down the same incline, you could test the effect of the variable:
    1. solid or hollow
    2. length of rod
    3. material of the rod
    4. angle of the incline
    5. none of the above

The item response pattern is:

Group		Foil Number 1	2	3	4	5	Omitted
Upper	27%	13%	13%	40%	7%	27%	0%
Middle	46%	27%	15%	27%	12%	4%	0%
Lower	27%	33%	7%	27%	0%	13%	20%
Total		25%	13%	30%	9%	16%	7%

Index of item discrimination .14
Index of item difficulty 84
The correct response is foil 5.

The student is asked to recognize the fact that there is more than one variable which could affect the time that it took the rods to roll down the incline and therefore, foil 5, none of the above, is the correct response. This apparently is too abstract for this age student.

Item 44

This item deals with information which is presented in a data table as well as in graphic form. The information represents the results of an experiment where students recorded the time that it took saccharin tablets to dissolve in water of different temperatures. The item read:

44. Did this group of students change the variable of time?

1. yes
2. no

The response pattern for this item was:

Group		Foil Number 1	2	Omitted
Upper	27%	60%	40%	0%
Middle	46%	62%	38%	0%
Lower	27%	33%	47%	20%
Total		54%	41%	5%

Index of item discrimination .07 negative
Index of item difficulty 59
The correct response is foil 2.

The author is unable to explain the reason for the failure of this item.

There were only three items in the process of interpreting data that were rejected and all three dealt with graphs. They were items 77, 11 and 12.

Item 77

Item 77 presents four graphs of supposedly the same data presented in a Table and asks the students to determine which is the best graph of that data. This item should be omitted as the author is unable to explain or suggest reasons for its failure. The item response pattern was:

Group		Foil Number 1	2	3	4	Omitted
Upper	27%	27%	13%	0%	53%	7%
Middle	46%	19%	12%	8%	58%	4%
Lower	27%	20%	13%	13%	40%	13%
Total		21%	13%	7%	52%	7%

Index of item discrimination .13
Index of item difficulty 48
The correct response is foil 4.

Items 11 and 12

Both of these items refer to a graph which showed a relationship between the amount of weight hung from the end of a stick and the height the end of the stick is from the floor. To answer these questions, the students were required to write in a numerical answer which represents the weight of the object causing the particular amount of bend. The students' responses, hereafter referred to as R, were then transcribed by this author to the machine scoring answer sheets in accordance with the following code. Foils for question 11 were:

1. $R \leq 25$ grams
2. 25 grams $< R \leq 50$ grams
3. 50 grams $< R < 100$ grams
4. 100 grams $\leq R < 200$ grams
5. $R \geq 200$ grams

The response pattern for this item was:

Group		Foil Number 1	2	3	4	5	Omitted
Upper	27%	7%	7%	67%	0%	13%	7%
Middle	46%	19%	4%	62%	4%	4%	8%
Lower	27%	27%	0%	53%	0%	7%	13%
Total		18%	4%	61%	2%	7%	9%

Index of item discrimination .14
Index of item difficulty 40
The correct response is foil 3.

Foils for question 12 were:

1. $R \leq 250$ grams
2. 250grams $< R \leq 375$ grams
3. 375 grams $< R < 425$ grams
4. 425 grams $\leq R < 525$ grams
5. $R \geq 525$ grams

The response pattern for this item was:

Group		Foil Number 1	2	3	4	5	Omitted
Upper	27%	0%	0%	27%	33%	33%	7%
Middle	46%	23%	0%	23%	8%	38%	8%
Lower	27%	7%	0%	20%	7%	53%	13%
Total		13%	0%	23%	14%	41%	9%

Index of item discrimination .07
Index of item difficulty 77
The correct response is foil 3.

Since these two items were numerical fill in, the author is at a loss to give suggestions for revisions which might improve the index of item discrimination. Therefore, it is suggested that these items be omitted.

After reading the last three items which were omitted, the reader may wonder about the relative success of graph-related items. In this study, there were fourteen items which were related to graphs. All except three received an index of item discrimination acceptable for retention.

Acceptable Items

We will now turn our attention to the acceptable items. It should be remembered that those items which had an index of item discrimination of .20 or greater were retained as acceptable items. Data concerning the frequency distribution and the indices for each item will be presented in the tables and graphs.

Table 7

Frequency Distribution from "A Group Test of Four Processes" for the process of Controlling Variable.
Acceptable Items Only
(Possible Score of 13)

Raw Score	Frequency	Cumulative Frequency	Percentile Rank
12	3	56	97
11	3	53	92
10	6	50	84
9	6	44	73
8	5	38	63
7	8	33	52
6	6	25	39
5	5	19	29
4	4	14	21
3	4	10	14
2	4	6	7
0	2	2	1

Range 0-12
Mean 6.7
Standard Deviation 3.1

Table 8

Frequency Distribution from "A Group Test of Four Processes" for the Process of Interpreting Data. Acceptable Items Only (Possible Score of 22)			
Raw Score	Frequency	Cumulative Frequency	Percentile Rank
20	3	56	97
19	3	53	92
18	3	50	87
17	1	47	83
16	3	46	79
15	7	43	71
14	5	36	60
13	3	31	53
12	3	28	47
11	4	25	41
10	4	21	34
9	1	17	29
8	4	16	25
7	3	12	19
6	3	9	13
5	2	6	9
4	2	4	7
3	1	2	3
1	1	1	1

Range 1-20
Mean 12.5
Standard Deviation 5.2

Tables 7 and 8 show a wide distribution which indicates the range of the students' abilities to perform on the pencil and paper test. Figures 3 and 4 may help the reader visualize these distributions.

Frequency Distribution from "A Group Test of Four Processes" for the Process of Controlling Variables. Acceptable Items Only. (Possible Score of 13)

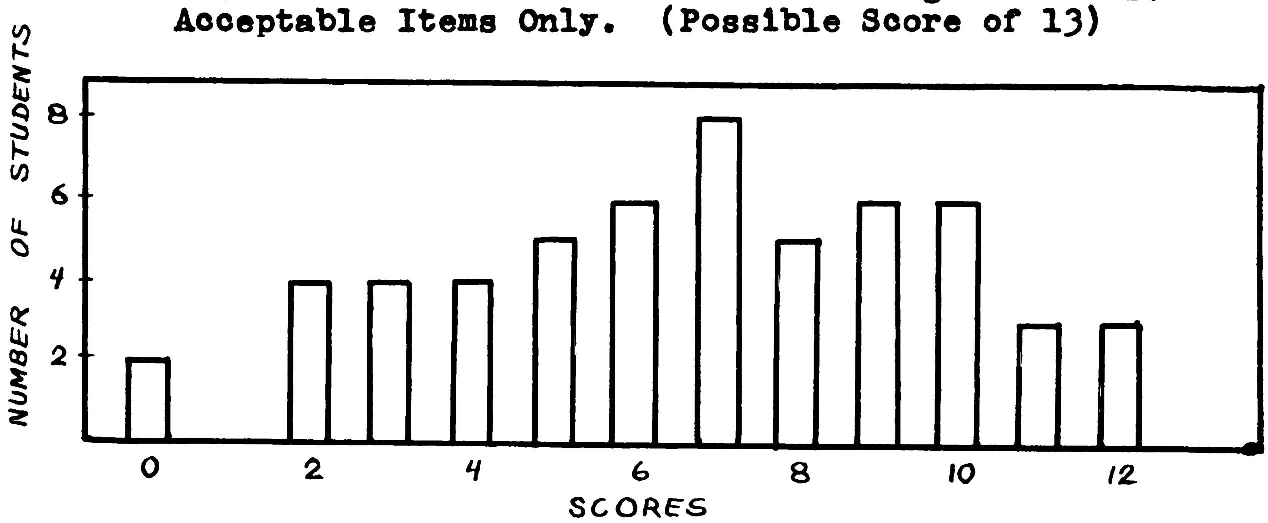


Figure 3

Frequency Distribution from "A Group Test of Four Processes" for the Process of Interpreting Data. Acceptable Items Only. (Possible Score of 22)

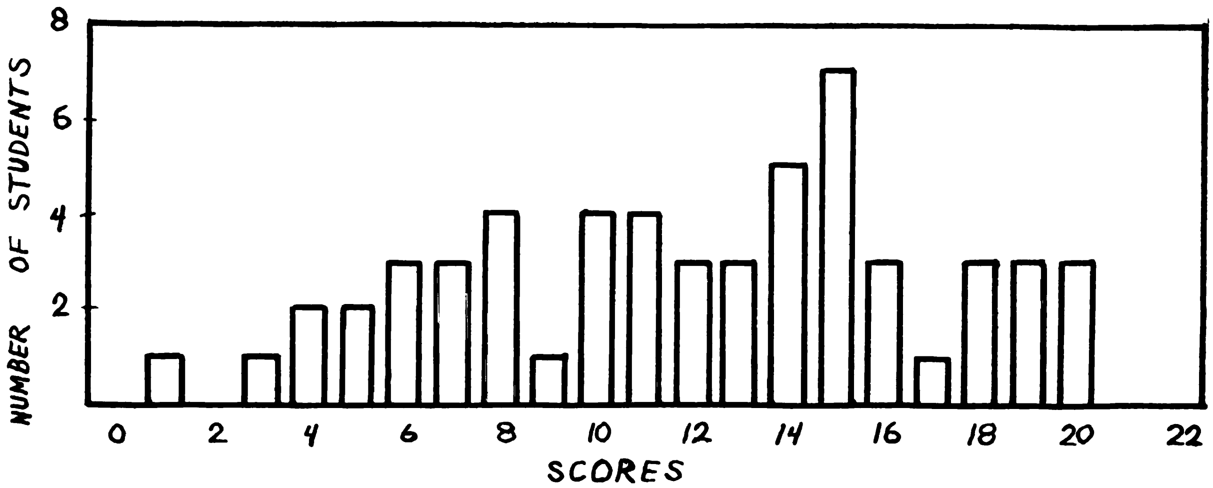


Figure 4

Both graphs, Figures 3 and 4, reveal similar distributions in that they are both slightly positively skewed. The reader is reminded that this is similar to the distributions of the criterion measure, the "Individual Competency Measures".

Tables 9 and 10 present the indices of item discrimination and item difficulty for each of the acceptable items for the processes of controlling variables and interpreting data.

Table 9

Acceptable Items from "A Group Test of Four Processes" for the Process of Controlling Variables

Item Number	Index of Discrimination	Index of Difficulty
25	.40	63
27	.34	29
29	.40	60
30	.34	54
31	.46	67
32	.47	38
34	.47	59
35	.40	35
36	.40	44
63	.53	43
64	.66	39
65	.46	46
78	.33	50
Means	.44	48

Table 10

Acceptable Items from "A Group Test of Four Processes" for the Process of Interpreting Data		
Item Number	Index of Discrimination	Index of Difficulty
3	.26	30
4	.20	71
5	.20	64
6	.40	42
13	.60	48
23	.26	68
26	.33	81
37	.53	57
38	.33	30
39	.20	36
40	.20	60
45	.33	40
46	.60	35
47	.46	29
48	.60	64
49	.20	57
66	.27	48
67	.26	37
68	.60	35
73	.53	21
74	.47	16
75	.40	36
Means	.37	46

Tables 9 and 10 reveals that all the acceptable items were contributing, but doing so with a variety of degrees of difficulty. The mean index of item difficulty for both processes are very close to the desired index of test items, .50. The complete item analysis data for each of the acceptable items in "A Group Test of Four Processes" can be found in Appendices B and C.

Items Presented Visually and Non-Visually

It was not the purpose of this study to determine the merits of items presented visually and non-visually, but a cursory look at the data is warranted. Of the 79 items prepared for "A Group Test of Four Processes", 40 were presented visually, with the aid of an oral script and a series of colored 35 mm slides. The other 39 items were presented in writing without oral communication and without the aid of colored slides. The following table presents the data comparing the indices of item discrimination and difficulty for all items presented visually and non-visually.[61]

---

[61]Part of this data was taken from the data presented by Darrel Wayne Fyffe in his thesis. op. cit.

Table 11

Comparison of Indices of Item Discrimination and Difficulty for the Items Presented Visually and Non-Visually

Item Discrimination				Item Difficulty			
Visual		Non-Visual		Visual		Non-Visual	
Score	f	Score	f	Score	f	Score	f
60	1	66	1	92	2	82	1
53	1	60	3	90	1	64	4
47	2	54	1	84	1	62	1
46	1	53	3	81	1	59	1
40	6	47	1	77	1	57	2
34	2	46	5	76	1	54	1
33	3	40	4	72	1	53	2
27	1	33	3	71	1	50	2
26	3	27	4	68	3	48	3
20	6	26	2	67	1	46	2
14	3	20	3	64	1	45	1
13	3	13	3	63	1	43	4
7	2	0	2	60	2	41	2
6	1	-07	2	59	1	40	2
0	4	-20	1	58	1	39	1
-20	1	-27	1	57	1	37	1
N 40		39		54	1	36	1
Mean 24.2		30.6		51	1	35	2
				49	1	29	2
				48	3	28	1
				44	1	21	2
				42	2	16	1
				40	1		

Table 11 (cont'd.)

38	1		
36	1		
35	1		
32	1		
30	2		
29	1		
21	1		
7	1		
N	40	39	
Mean	54.7	45.7	

To better understand whether there was a significant difference in those items presented visually and those presented non-visually, a "t-test" was run.

Table 12

"t-test" for the Comparison of Indices of Stem Discrimination and Difficulty for the Stems Presented Visually and Non-Visually	
Critical Region: $R \geq \lvert t \rvert$ 1.99	a=0.05 df=77
Index of Items:	Calculated "t"
Discrimination	1.51
Difficulty	2.35

The calculated "t" for the index of item discrimination indicates that there is no significant difference in the indices of item discrimination for the items presented visually and non-visually. On the other hand, the "t" for the index of item difficulty was significant at the .05 level. The calculated "t" indicates that those items which were presented visually had a higher mean index of item difficulty than those which were presented non-visually. Perhaps the format of presenting items visually could have been new to the students. A more plausible explanation for the significant difference is that the visual format was chosen because the items were judged to be more difficult and to require more verbage. Even though a visual format was chosen, the items probably remained more difficult.

Correlations

The purpose of this study was to develop group test items that would measure the same skills as those required in the selected "Individual Competency Measures" for the processes of controlling variables and interpreting data. The task was to produce items that possess a reasonable index of discrimination based on the criterion measure. If they also have a high index of correlation with the student's ability to perform (or not perform) the related processes, the assumption is made that both the items and the competency tasks are measuring the same thing.

Table 13 presents the raw data for each student's score on the individual and group tests for the processes of controlling variables and interpreting data.

Table 13

Scores Received on the "Individual Competency Measures" (ICM) and "A Group Test of Four Processes" (GTFP) for the Process of Controlling Variables (CV) and Interpreting Data (ID)

Student Number	CV ICM	CV GTFP	ID ICM	ID GTFP	Student Number	CV ICM	CV GTFP	ID ICM	ID GTFP
1	12	10	14	20	34	11	9	9	14
2	12	5	12	10	35	4	2	7	5
3	10	9	11	15	37	10	7	6	8
4	8	4	7	12	38	8	10	6	15
5	11	10	14	20	40	3	5	4	4
8	0	2	3	7	41	1	4	3	8
9	7	7	11	15	43	9	10	10	19
10	7	8	7	10	44	1	7	3	6
11	10	9	10	14	45	7	4	8	11
12	10	7	10	16	46	8	7	12	15
13	9	9	9	16	47	5	4	7	6
14	5	6	14	10	48	9	12	10	20
15	9	5	11	13	49	12	7	5	11
16	10	7	10	12	51	8	8	5	11
17	8	3	11	13	52	10	8	8	17
18	5	5	9	5	53	11	11	10	11
20	6	3	9	7	54	6	3	10	10
21	3	3	8	8	55	11	9	12	13
22	7	6	8	15	56	6	7	11	12
23	5	5	6	7	57	12	12	14	14
24	5	6	1	6	59	11	11	9	14
25	7	11	10	18	61	11	12	14	19

Table 13 (cont'd.)

26	9	10	13	18	63	6	6	11	15
27	1	0	5	4	65	11	10	13	18
28	6	2	1	8	66	8	8	12	19
30	4	0	10	3	67	7	8	10	15
31	4	2	4	1	68	6	6	8	9
33	6	6	10	14	69	8	9	9	16

Some student numbers will be missing (example, numbers 6 and 7) for various reasons including being absent from either or both instruments and not attending Kinawa Middle School as a sixth grader.

For the processes of controlling variables and interpreting data, the analysis reveals a substantial correlation between the individual performance test and those retained items (item discrimination of .20 or greater) that were designed to measure the processes of controlling variables and interpreting data.

Table 14

Pearson-Product Moment Correlation Coefficient for the Processes of Controlling Variables (CV) and Interpreting Data (ID)

	Individual Competency Measures		
		CV	ID
A Group Test of Four Processes	CV	0.70	
	ID		0.66

In the testing of significance for the Pearson-r, the author referred to Table 6 in Downie and Heath[62] and

[62]N.M. Downie and R.W. Heath, Basic Statistical Methods, 2nd edition, New York: Harper & Row, Publishers, 1965, p. 306.

found that the correlation was significant at better than the 0.001 level.

While correlation statistics always carry with it some degree of inferential risk, this level of significance implies a high degree of confidence that the items in the group test are measuring substantially the same thing as the "Individual Competency Measures". In other words, the items would be good predictors of the scores students would receive when they are administered the "Individual Competency Measure" (the criterion measure).

Other Correlations

It seems logical that the ability to use the processes of science may be interdependent among the various processes. That is to say, ability to use one of the processes may be a predictor for the ability to use the others. Since the group test contained items representing four different processes, controlling variables, interpreting data, formulating hypotheses, defining operationally, it was considered a pursuit of interest to investigate the interdependency among the processes. Data for the processes of formulating hypotheses and defining operationally can be obtained from Fyffe.

The correlations for the various combinations of processes are shown below.

Table 15

Correlation Coefficients of "Individual Competency Measure" Scores and Group Test Scores

		Individual Competency Measures				Group Test Items			
		CV[63]	ID	FH	DO	CV	ID	FH	DO
ICM	CV	1							
	ID	.601	1						
	FH	.540	.583	1					
	DO	.573	.536	.430	1				
GTI	CV	.705	.503	.598	.483	1			
	ID	.703	.660	.414	.502	.779	1		
	FH	.656	.525	.535	.434	.618	.561	1	
	DO	.664	.537	.425	.565	.786	.728	.662	1

Table 16[64]

Values of Pearson-Product Moment Correlation Coefficient Required for Specific Levels of Significance (54 degrees of freedom)

Level of Significance	Correlation Required
.05	.265
.01	.343
.001	.430

Table 18 indicates that any correlation coefficient, from Table 17, that is equal to or greater than .430 is significant at the .001 level. A look at Table 17 reveals that out of the thirty-six possible correlations, only two

---

[63]CV = Controlling Variables, ID = Interpreting Data, FH = Formulating Hypotheses, and DO = Defining Operationally.

[64]Downie and Heath, op. cit.

were not significant at the .001 level. The two correlation coefficients, the "Individual Competency Measure" score for formulating hypotheses with the group test item scores for interpreting data and defining operationally, are, however, significant at the .01 level. This data gives a strong indication, significant at the .001 level with the exception of the two mentioned before, of the interdependency among the processes of controlling variables, interpreting data, formulating hypotheses and defining operationally. The evidence suggests, for example, that it is possible to predict a student's ability to control variables, from his score on any of the processes. The title, "integrated processes", appears to be a very accurate one.

Summary

In this chapter, the data is presented relative to the development of objective test items for the processes of controlling variables and interpreting data. Thirteen of the original eighteen items for the process of controlling variables were found to have a satisfactory index of item discrimination (.20 or greater) while twenty-two of the twenty-five items were satisfactory for the process of interpreting data. Those items which were not satisfactory were reviewed by this author and suggestions for revisions made where they seemed appropriate.

This study also indicated ambiguity in the significance of those items which were presented visually and those which were presented non-visually. Even though there was no significant difference (.05 level) for the index of item discrimination, a significant difference was apparent in the index of item difficulty. Those items which were presented visually were more difficult than the items which were not presented visually. This could possibly be explained by the fact that the visual format was chosen because the situation was more difficult and required more verbage. Even though a visual format was used, it was probably still a more difficult item.

Correlation coefficients have been reported which indicates the relationship between the scores on the "Individual Competency Measures" and the scores on the group test items for the two processes of controlling variables and interpreting data. The data indicates a correlation significant, in both cases, at the .001 level. Other correlation data indicates that there is great interdependency among the processes of controlling variables, interpreting data, formulating hypotheses and defining operationally. Thirty-four of the thirty-six pairs reported had correlations significant at the .001 level with the other two significant at the .01 level.

CHAPTER V

SUMMARY, CONCLUSIONS, FUTURE RESEARCH

Summary

With the increasing emphasis in the instruction in the processes of science, there is a need to develop test items that test the acquisition of these processes in an efficient way.

This study was an attempt to develop objective test items which will assess seventh grade students' ability to use the processes of controlling variables and interpreting data. The procedure followed in constructing and determining acceptable items used an external measure as the criterion for validation. The criteria for success of individual test items were two of the "Individual Competency Measures" for the process of controlling variables (Controlling Variables 7 and 9), and three of the "Individual Competency Measures" for the process of interpreting data (Interpreting Data 9, 12 and 13), all from Part 6 of Science--A Process Approach. Each student was tested individually for approximately two hours on the selected "Competency Measures" and was scored on his ability to perform the behavioral objectives specified in the measures.

After the individual testing, the author, in cooperation with Darrel W. Fyffe, wrote and assembled a 79 item group test, "A Group Test of Four Processes". The items were designed to measure the processes of controlling variables, interpreting data, formulating hypotheses and define operationally. The latter two processes were of interest to Fyffe. The group test was administered to a sample of 56 seventh grade students who previously had Science--A Process Approach.

Item analysis data were obtained which reported the index of item discrimination, index of item difficulty, and gave the item response pattern. The acceptable index of item discrimination was chosen at .20 because the item validity was based on an external independent performance measure. The analysis of this study showed that 13 of the original 18 items on the process of controlling variables and 22 of the original 25 items on the process of interpreting data achieved the predetermined index of item discrimination level of .20 or greater.

Another analysis was performed to show the correlation coefficients between the scores on the individual test and the scores on the group items with high discrimination (.20 or greater). The Pearson product moment correlation coefficient showed a highly significant correlation (significant at the .001 level) between the individual and group scores for each of the two processes.

Conclusions

The findings of this study do, in fact, show that the procedure of using an independent measure as a criterion for producing objective test items is valid for the processes of controlling variables and interpreting data. Through this study, 13 valid test items for the process of controlling variables and 22 valid test items for the process of interpreting data were developed.

The data reported in this study also gives evidence that there is a high degree of interdependency among the four processes of controlling variables, interpreting data, formulating hypothesis and defining operationally, since the scores on any one of the processes seem to be predictors of the scores on the other processes.

Implications for Future Research

The items produced in this study carry both criterion-related validity and high indices of item discrimination, but they, in no way, represent a finished test over these processes of science. The test items developed in this study, and/or other items which meet the requirements of criterion-related validity, need to be constructed into a test which measures process skills.

Besides the implication of developing a test for the processes of controlling variables and interpreting data, this research suggests other studies which would be of interest.

1. Once a test is constructed with items possessing criterion-related validity, a correlation study needs to be conducted using existing process tests which report only content validity. Since this method of test item construction is very time consuming, there is a need to find out if content validity is sufficient.
2. This study also shows that the "Individual Competency Measures" for the processes of controlling variables and interpreting data are substantially correlated with the "Individual Competency Measures" for the processes of controlling variables, interpreting data, formulating hypotheses and defining operationally. This correlation data, ranging from .62 to .79, is significant at the .001 level. This implies that the integrated processes, especially the processes of controlling variables, interpreting data, formulating hypotheses and defining operationally, are in fact composed of common skills. The best guess of what these common skills are, seems to be the basic processes of science which are defined in the science program, <u>Science--A Process Approach</u>. To test this, a class of students could be given instruction in a subset of the basic processes, starting with observing

and working up the hierarchial ladder, to determine which, if any, of the subsets can be considered prime contributors to the ability to use the integrated processes. Another possibility could be that performance on the process test is substantially measuring some more basic attribute such as I.Q. or reading level. The interdependency suggests further investigation into those questions once a completed test is available.

3. This study also implies that by following the procedure outlined in this study, valid, objective test items could be developed for all the processes of science, the basic processes as well as the integrated processes.

4. This procedure could also be extended to areas where test item development has been thought to be impractical. These areas, where evaluation has been based on observation of individual behaviors such as attitudes and interests are suitable for using this procedure to develop test items. The first task would be to specify in behavioral terms what is to be tested. Once this is determined, then the procedure described in this study could be utilized to develop objective test items with criterion-related validity.

5. This research also indicates that a study needs to be conducted which would investigate differences on the success of items that are presented visually and non-visually.
6. The availability of appropriate means of evaluating the processes of science should result in greater concern about evaluating the success of science programs which claim to emphasize the processes of science.
7. The format used in administering the first part of "A Group Test of Four Processes" (the visual part) raises another implication which should be investigated. Students were not allowed to work at their own rate. The pace was set by the slower students. Could this restriction cause the faster students to become bored or less interested in the task at hand? This could be handled by substituting, for the colored slides, pictures printed in the test booklet.
8. The science content used in process tests could have an effect on the success of the students making correct responses. Research needs to be conducted on the content effect as well as the transfer of process across different content areas.

9. Any test which requires the student to do their own reading is always subject to the question of the reading level of the student compared to the reading level of the test. Would other test formats, such as taping, produce a more accurate measure of a student's ability to perform the processes?
10. The population of students chosen for this study were familiar with the vocabulary used in Science--A Process Approach. They also belong to an upper middle class, suburban community. This implies that a process test which uses these items could be designed for only a portion of the educational population. An investigation relating to the use of a process test with students having different cultural and economical backgrounds needs to be conducted.
11. The teaching of scientific processes, like everything else, is subject to the methodological philosophies of the teachers involved. In most cases, teachers are not able to completely adopt someone else's methodology. When a school district adopts a program emphasizing the processes of science, the teachers modify the suggested methodology to fit within their own structure. Questions such as, "Are processes

best taught with an open-discovery hands-on approach, teacher-directed hands-on approach or with the 'lecture' approach?" need to be investigated. With the establishment of a valid science process test, it would be possible to evaluate the effect of different methods of teaching on process attainment.

12. A measure of success on a science process test could also be of value to a vocational counselor. Is the ability to use science processes a predictor of success in subsequent science courses or in the field of science?

Finally, there is an implicit assumption by educators that process skills are generalizable. This study was concerned with the scientific processes of controlling variables and interpreting data. There is a need to investigate the transfer of these processes to other academic areas. For example, the processes used in this study are applicable to the field of social science. Just exploring the area of social science, many questions could be investigated. Are the definitions for the processes of controlling variables and interpreting data the same in science as they are in social science? Is success in the science processes of controlling variables and interpreting data a predictor of success in the same social science processes? Is it possible to develop a single process test which will assess the processes for both academic fields? Is it necessary to teach the processes in both subject matter areas?

BIBLIOGRAPHY

BIBLIOGRAPHY

Books

American Association for the Advancement of Science, An Evaluation Model and its Application. Second Report. Washington D.C., 1968.

________. Science--A Process Approach Commentary for Teachers. AAAS Misc. Publication 68-7, 1968.

________. Science--A Process Approach, Part A, New York: XEROX Education Division, 1967.

________. Science--A Process Approach, Part B, New York: XEROX Education Division, 1967.

________. Science--A Process Approach, Part Six, Fourth Experimental Edition, AAAS Misc. Publication 67-11, 1967.

________. The Psychological Bases of Science--A Process Approach, AAAS Misc. Publication 65-8, 1965.

Biological Science Curriculum Study. Processes of Science Test, New York: The Psychological Corporation, 1962.

Bloom, Benjamin S. Taxonomy of Educational Objectives, Handbook 1: Cognitive Domain, New York: David McKay Company, Inc., 1956.

Bruner, Jerome S. The Process of Education, New York: Random House, Inc., 1960.

Buros, Oscar K. (ed.). The Sixth Mental Measurements Yearbook, New Jersey: The Gryphon Press, 1965.

Campbell, Donald F. and Julian C. Stantey. Experimental and Quasi-Experimental Designs for Research, Chicago: Rand McNally and Company, 1963.

Downie, N.M. and R.W. Heath. Basic Statistical Methods, Second edition, New York: Harper & Row, Publishers, 1965.

Ebel, Robert L. Measuring Educational Achievement, New Jersey: Prentice-Hall, Inc., 1965.

Gagne, Robert M. The Conditions of Learning, New York: Holt, Rinehart and Winston, Inc., 1965.

Horrocks, John E. and Thelma I. Schoonover. Measurement for Teachers, Ohio: Charles E. Merrill Publishing Company, 1968.

Hurd, P. and J. Gallagher. New Directions in Elementary Science Teaching, California: Wadsworth Publishing Co., Inc., 1969.

Mehrens, William A. and Irvan J. Lehmann. Standardized Tests in Education, Holt, Rinehart and Winston, Inc., 1969.

Piaget, Jean. Six Psychological Studies, New York: Random House, Inc., 1967.

Skinner, B.F. The Technology of Teaching, New York: A. Century-Crofts, 1968.

Sund, Robert B. and Leslie W. Trowbridge. Teaching Science by Inquiry in the Secondary Schools, Ohio: Charles E. Merrill Books, Inc., 1967.

Periodicals

Burns, Richard W. and Gary D. Brooks. "What are Educational Processes?" The Science Teachers, XXXVII (February 1970), 27-28.

Cole, Henery P. "Process Curricula and Creativity Development," Journal of Creative Behavior, III (Fall 1969), 244.

Freud, H. and N.D. Cheronis. "Retention in the Physical Science Survey Course," Journal of Chemical Education, XVII, 1940, 289-293.

Frutchey, F.P. "Retention in High School Chemistry," Journal of Higher Education, VIII, 1937, 217-218.

Gagne, Robert M. "Contributions to Human Development," Psychological Review, LXXV, No. 3, 1968, 177-191.

________. "Learning Hierarchies," Presidential Address, Division 15, American Psychological Association, (San Francisco, August 1968).

Tyler, Ralph W. "Permanence of Learning," Constructing Achievement Tests, Columbus: Ohio State University, 1934.

Welch, Wayne W. "Curricular Decisions: How can Evaluation Assist Science Teachers?" The Science Teacher, XXXV (November 1968), 22-25.

Unpublished Works

Beard, Jean. "Group Achievement Tests Developed for Two Basic Processes of AAAS Science--A Process Approach." Unpublished dissertation, Oregon State University, 1970.

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

Sweetser, Evan. "Science Process Test for Elementary School Teachers," Michigan State University.

Tannenbaum, Robert Sher. "The Development of the Test of Science Processes." Unpublished dissertation, Teachers College, Columbia University, 1968.

APPENDICES

APPENDIX A

SCRIPT, SLIDES AND ITEMS

USED IN "A GROUP

TEST OF FOUR PROCESSES"

APPENDIX A

SCRIPT, SLIDES AND ITEMS
USED IN "A GROUP
TEST OF FOUR PROCESSES"

The test items developed in this study follow. Also included is the script that was read to the students during administration of the group test items. The photographs are reproductions of the color slides which were projected.

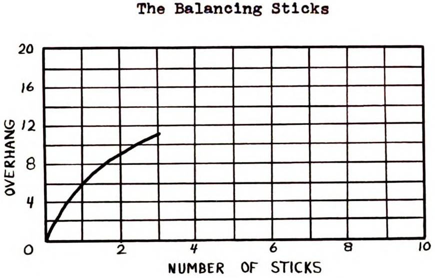
The following slides show the maximum overhang the class got by using 1, 2, and 3 sticks. With one stick the overhang was 6 units. Show slide 4.

With two sticks they got a 9 unit overhang. <u>Show slide 5</u>.

With three sticks the overhang was 11 units. <u>Show slide 6</u>.

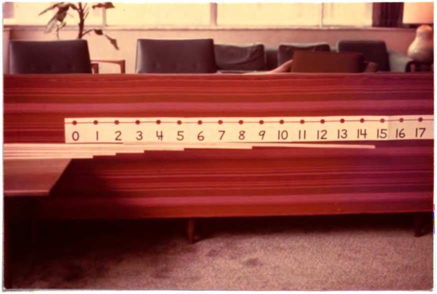
Now answer question 3 on page 1.

3. From the graph below, what is the least number of sticks that you would need to get an overhang of 12 units or more?



STOP. WAIT FOR INSTRUCTIONS.

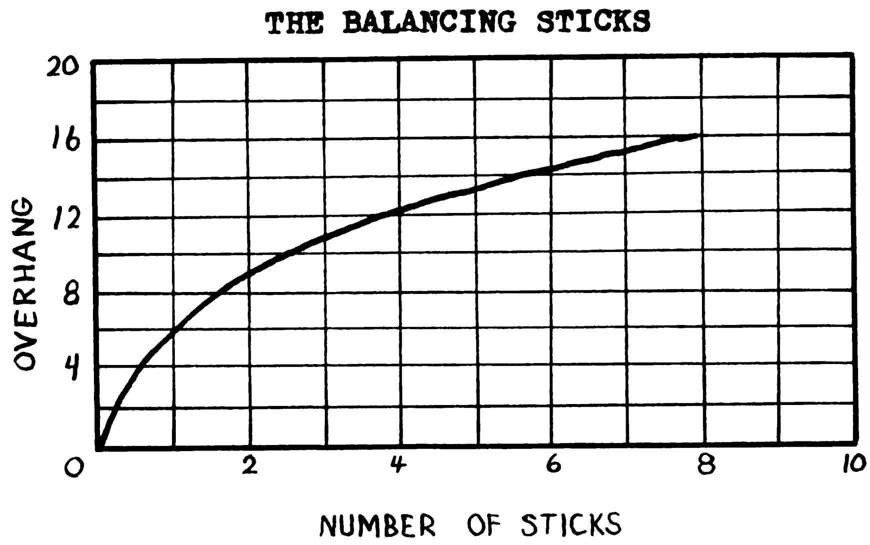
With six sticks the overhang was 14 1/2 units. Show slide 7.



With eight sticks they got an overhang of 16 units.

Show slide 8.

Turn to page 2 and answer all three questions.



Using the graph above, answer the following questions.

4. What would be the maximum overhang you could get by using 5 sticks?

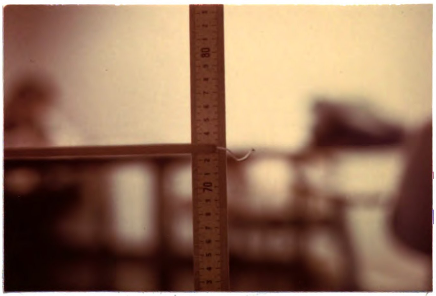
5. What is the fewest number of sticks you could use and get an overhang of at least 15 units?

6. How much of an overhang could you get by using 9 sticks?

STOP. WAIT FOR INSTRUCTIONS.

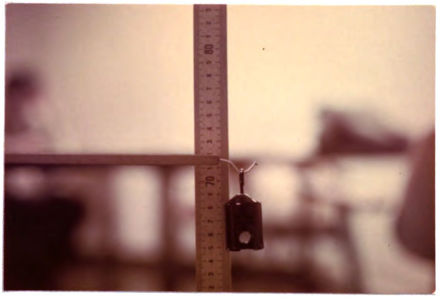
By using this set-up, (Show slide 13.)

a science class found that with no weight suspended from the end of the stick, (Show slide 14.)



the bottom of the stick was 72 1/2 cms. from the floor.

When 100 grams was hung on the hook at the end of the stick, (Show slide 15.)



it bent so that the bottom of the stick was 71.2 cms. from the floor. With 200 grams (Show slide 16.)

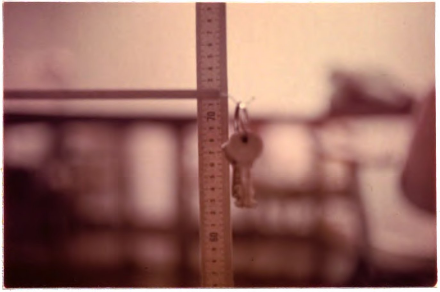
it was 69.9 cms. from the floor.

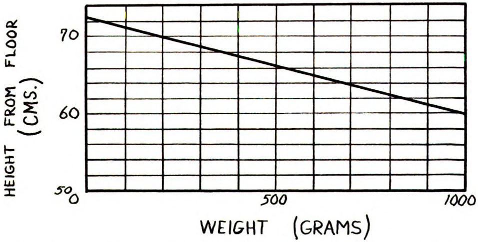
With 500 grams (<u>Show slide 17.</u>)

it was 66.2 cms. from the floor and with 1000 grams (<u>Show slide 18</u>.)

it was 60.1 cms. from the floor. Turn to page 6.

This slide is for question 11. (Show slide 19.)





Using the above graph, answer the following question.

11. The keys cause the stick to bend so it is 71.4 centimeters from the floor. What is the weight of the keys?

STOP. WAIT FOR THE NEXT SLIDE.

This slide is for question 12. (Show slide 20.)

12. The series of four weights bend the stick to 67.8 centimeters from the floor. What is their weight?

GO ON TO THE NEXT QUESTION.

13. If 600 grams were suspended from the end of the stick, what would be the height from the floor?

STOP. WAIT FOR INSTRUCTIONS.

A student (Show slide 27.)

wanted to see if a beaker that contained a saltwater solution would freeze in different lengths of time when it was placed in freezers of different temperatures. (Show slide 28.)

He found that at -10 degrees it took 60 minutes for the saltwater to freeze. (Show slide 29.)

At -20 degrees it took 45 minutes. (Show slide 30.)

At -30 degrees it took 35 minutes for the saltwater to freeze. (Show slide 31.)

From this information (Show slide 32.)

Answer questions 23, 24, and 25 on page 9.

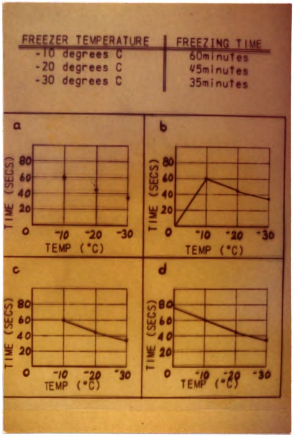
23. From this information, we can determine that:
    1. For every 10 degrees drop in temperature, the time decreases the same amount.
    2. For every 10 degrees drop in temperature, the time increases the same amount.
    3. As the temperature increases, the length of time to freeze the solution increases.
    4. A beaker containing a salt solution and placed in a freezer at -25 degrees Celsius, would freeze in 40 minutes.

24. Which of the following variables were changed in this investigation?
    1. amount of salt in the solution.
    2. temperature of the freezer.
    3. time required for the mixture to freeze.
    4. size of the container.

25. Which of the variables in the above question were kept constant.
    1. amount of salt
    2. amount of salt and temperature
    3. amount of salt and time
    4. amount of salt and size
    5. amount of salt, time and size

<u>STOP. WAIT FOR THE NEXT SLIDE.</u>

This slide is for question 26. (Show slide 33.)

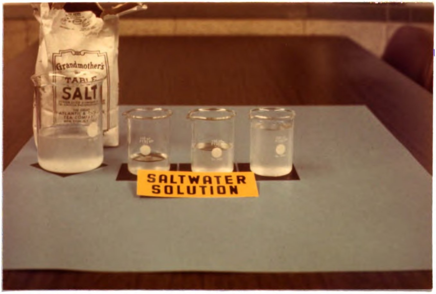


26. Which is the best graph of the information?
    1. a
    2. b
    3. c
    4. d
    5. none of these

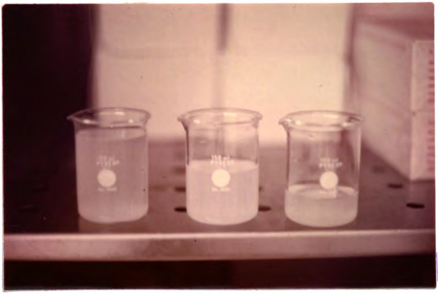
STOP. WAIT FOR INSTRUCTIONS.

In another investigation, (<u>Show slide 34</u>.)

the class used 3 glass beakers all the same size. To each beaker, (<u>Show slide 35</u>.)



different amounts of a saltwater solution were poured from the larger beaker. The 3 glass beakers (Show slide 36.)



were then placed in a freezer compartment. Now answer question 27.

27. Through this investigation the students will be able to determine if the time required to freeze the solutions depends on the:
    1. temperature of the freezer compartment.
    2. size of the container.
    3. material from which the container is made.
    4. amount of solution in the container.

STOP. WAIT FOR THE NEXT SLIDE.

Turn t

28. E
c

29.

STOP

(Show slide 37.)

Turn to page 10 and answer questions 28 and 29.

28. By using this set-up, they could investigate the effect of different:
    1. amounts of salt.
    2. kinds of containers.
    3. sizes of containers.
    4. freezer temperatures.
    5. none of the above

29. Which of the following variables would you have to change to determine how long it takes the solution to freeze in different containers?
    1. amount of salt
    2. temperature of the freezer
    3. kind of container
    4. time for the solution to freeze

STOP. WAIT FOR THE NEXT SLIDE.

This slide is for question 30. (Show slide 38.)

30. A group of students wanted to test the hypothesis that an increase in the amount of salt will decrease the time to freeze the mixture. Which of the experiments on the slide would test this hypothesis?
    1. experiment 1
    2. experiment 2
    3. experiment 3
    4. experiment 4

STOP. WAIT FOR INSTRUCTIONS.

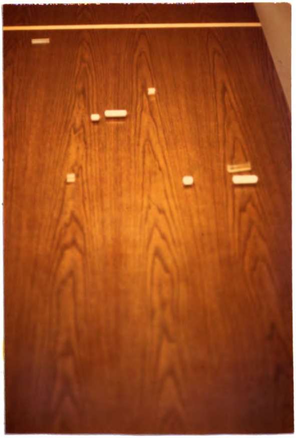
By

the clas

it takes

By raising one end of a table, (Show slide 39.)

the class was able to find out what affects the time that it takes rods to roll down an incline. (Show slide 40.)



T

shown her

Using th

turn to

The rods that were used in this experiment are shown here. (Show slide 47.)

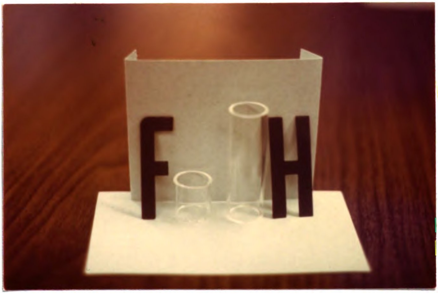
Using this slide (Show slide 41.)

turn to page 11 and answer question 31.

31. By rolling rods A and E down the same incline, you could test the effect of the variable:
    1. solid or hollow
    2. length of rod
    3. material of the rod
    4. angle of the incline
    5. none of the above

STOP. WAIT FOR THE NEXT SLIDE.

This slide is for question 32. (Show slide 42.)



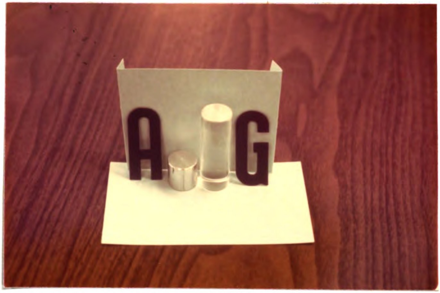
32. By rolling rods F and H down the same incline, you could test the effect of the variable:
    1. solid or hollow
    2. length of rod
    3. material of the rod
    4. angle of the incline
    5. none of the above

STOP. WAIT FOR THE NEXT SLIDE.

33. By
co

STOP.

This slide is for question 33. (Show slide 43.)



33. By rolling rods A and G down the same incline, you could test the effect of the variable:
    1. solid or hollow
    2. length of rod
    3. material of the rod
    4. angle of the incline
    5. none of the above

STOP. WAIT FOR THE NEXT SLIDE.

This sl

34. W
t
r

STOP.

This s

This slide is for question 34 on page 12. (Show slide 44.)

34. Which of the rods in the slide should be rolled down the same incline to determine whether or not a hollow rod rolls at a different rate than a solid rod.
    1. A and D
    2. A and F
    3. C and D
    4. C and H

STOP. WAIT FOR THE NEXT SLIDE.

This slide is for question 35. (Show slide 45.)

35. W
t
t

STOP.

This sl

36. W
i
t
v
t

STOP.

35. Which rod in the slide would you use to determine if the material of the rod affects the time the rod takes to roll down the incline?
    1. A and F
    2. A and H
    3. C and G
    4. C and H
    5. none of the above

STOP. WAIT FOR THE NEXT SLIDE.

This slide is for question 36. (Show slide 46.)

36. When some students rolled rods B and D down the same incline, they found that both rods took 10 seconds to travel the same distance. From just this information, which of the following variables does not affect the time?
    1. length of the rod
    2. hollow or solid
    3. material of the rod
    4. both 2 and 3
    5. none of the above

STOP. WAIT FOR INSTRUCTIONS.

These are the rods that were used in the experiment. (Show slide 47.)

Using the data table at the top of page 13, answer questions 37 through 42 on pages 13 and 14.

ROD	MATERIAL	LENGTH	TYPE	TIME
A	metal	2 cm	solid	5 secs
B	metal	2 cm	hollow	10 secs
C	metal	8 cm	solid	5 secs
D	metal	8 cm	hollow	10 secs
E	plastic	2 cm	solid	5 secs
F	plastic	2 cm	hollow	10 secs
G	plastic	8 cm	solid	5 secs
H	plastic	8 cm	hollow	10 secs

37. The material from which the rod is made affects the time for the rod to roll down the incline.
    1. true
    2. false
    3. cannot predict

38. Whether the rod is hollow or solid affects the time for the rod to roll down the incline.
    1. true
    2. false
    3. cannot predict

39. The length of the rod affects the time for the rod to roll down the incline.
    1. true
    2. false
    3. cannot predict

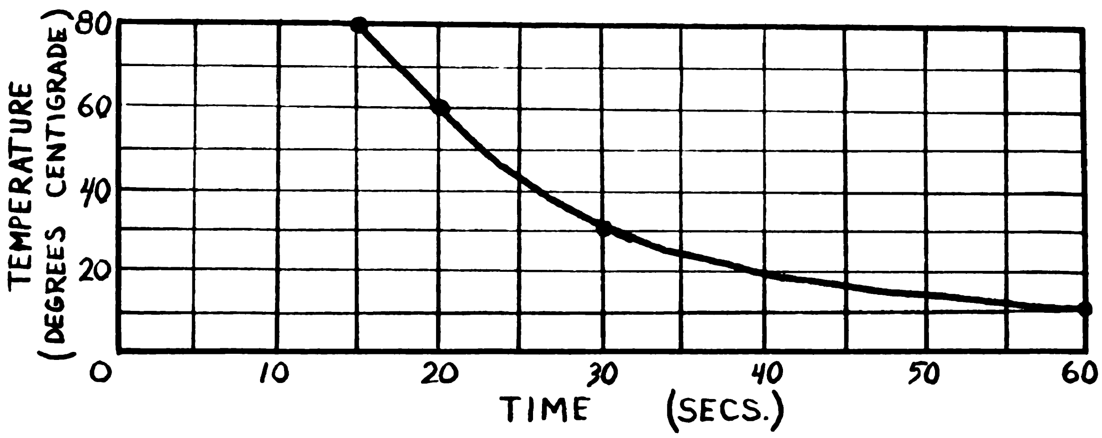
40. The angle of the incline affects the time for the rod to roll down the incline.
    1. true
    2. false
    3. cannot predict

STOP. WAIT FOR INSTRUCTIONS.

The following information is used for questions 43 to 50.

A group of students dissolved saccharin tablets in one cup of water. They measured the time required for one tablet to dissolve in water of different temperatures. The results of this experiment are listed below in the data table and pictured in the graph.

TEMPERATURE OF WATER (degrees Centigrade)	TIME TO DISSOLVE (seconds)
10 degrees	60
30 degrees	30
60 degrees	20
80 degrees	15



44. Did this group of students change the variable of time?
    1. yes
    2. no

45. From this information we can say the:
    1. dissolving time increases with increases in temperature.
    2. dissolving time decreases with increases in temperature.
    3. dissolving time decreases with decreases in temperature.
    4. temperature of the water has no effect.

46. From the graph, how long would it take for a saccharin tablet to dissolve at 20 degrees Centigrade?
    1. 20 seconds
    2. 40 seconds
    3. 45 seconds
    4. 60 seconds

47. If you want a saccharin tablet to dissolve in 25 seconds, you should use water at a temperature of:
    1. 20 degrees Centigrade
    2. 35 degrees Centigrade
    3. 40 degrees Centigrade
    4. 60 degrees Centigrade

48. Between which two points on the graph is the temperature changing most rapidly?
    1. 15 to 20 seconds
    2. 20 to 30 seconds
    3. 40 to 50 seconds
    4. 55 to 60 seconds

49. As the temperature changes from 10 degrees Centigrade to 60 degrees Centigrade the time needed for the saccharin tablet to dissolve:
    1. increases 40 seconds.
    2. increases 50 seconds.
    3. decreases 40 seconds.
    4. decreases 50 seconds.

50. Which of the following variables would you hold constant to test the hypothesis that the number of tablets dissolved in the water has no effect on the time needed to dissolve another tablet?
    1. amount of water
    2. number of saccharin tablets
    3. temperature of the water
    4. both amount and temperature of the water

The following information is used for questions 63 to 68.

With the help of the teacher, a science class set up an experiment where the students reacted to a light or a buzzer. Each student was presented with the stimulus (light, buzzer or both) and a timer started running. They would turn off the timer by flipping a switch and their reaction time was read from the timer. The class tried the six following experimental situations:

A. boys reacting to a light
B. boys reacting to a buzzer
C. girls reacting to a light
D. girls reacting to a buzzer
E. boys reacting to a light and a buzzer at the same time
F. girls reacting to a light and a buzzer at the same time

63. Which of the above experiments will be useful in deciding whether boys react more quickly to a light or a buzzer?
   1. E and F
   2. B and D
   3. A and C
   4. A and B

64. Which of the above experiments will be useful in deciding whether boys or girls react more quickly to a buzzer?
   1. E and F
   2. B and E
   3. B and D
   4. D and F

65. Which of the above experiments will be useful in deciding whether boys or girls react more quickly for either stimulus alone?
   1. A,B,C and D
   2. A,B,E and F
   3. A,C,E and F
   4. B,D,E and F

After the science class completed the six experiments, the following information was collected:

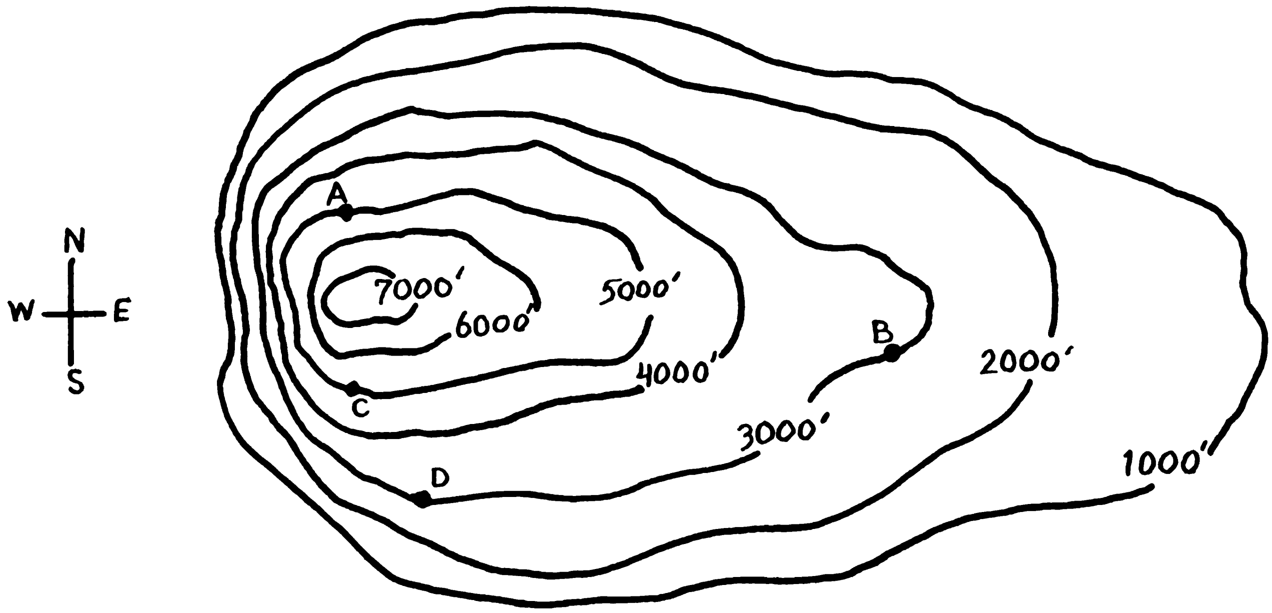
	STUDENTS TESTED	STIMULUS PRESENTED	AVERAGE TIME
A.	boys	light	0.17 seconds
B.	boys	buzzer	0.22 seconds
C.	girls	light	0.15 seconds
D.	girls	buzzer	0.19 seconds
E.	boys	both	0.14 seconds
F.	girls	both	0.23 seconds

66. Who reacts more quickly to a buzzer?
    1. Boys by .08 seconds.
    2. Boys by .03 seconds.
    3. Girls by .09 seconds.
    4. Girls by .03 seconds.

67. Did the boys react faster to a light than girls react to a buzzer?
    1. Yes, by .02 seconds.
    2. Yes, by .09 seconds.
    3. No, girls are faster by .05 seconds.
    4. No, girls are faster by .08 seconds.

68. Who reacts more quickly to a light and a buzzer at the same time?
    1. Boys by .02 seconds.
    2. Boys by .09 seconds.
    3. Girls by .09 seconds.
    4. Girls by .05 seconds.

Use the following contour map to answer the following three questions.



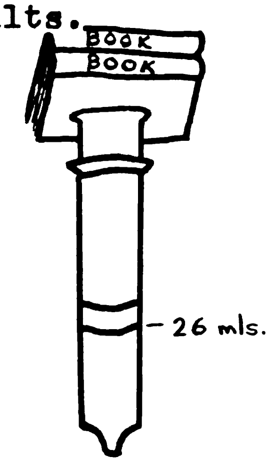
73. What is the height of point A?
    1. 9000'
    2. 6500'
    3. 6000'
    4. none of the above

74. What is the height of B?
    1. 2000'
    2. 2500'
    3. 3000'
    4. none of the above

75. This mountain has its steepest side on its:
    1. north side.
    2. south side.
    3. east side.
    4. west side.
    5. not enough information given to tell.

The following information is used for questions 77 and 78.

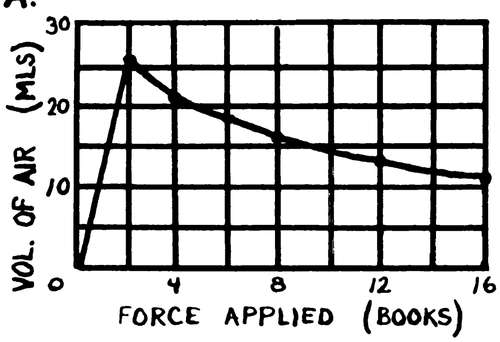
Books were placed on the plunger of a syringe that was filled with air. As books were added to the top of the plunger, the air inside the syringe was compressed more. Below are the results.



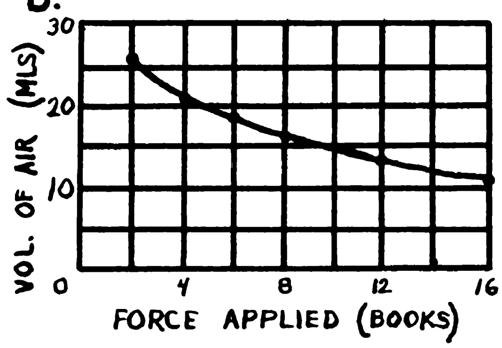
FORCE APPLIED (books)	VOLUME OF AIR (milliliters)
2	26
4	21
6	18
8	16
12	13
16	11

77. Which of the following is the best graph of the data?

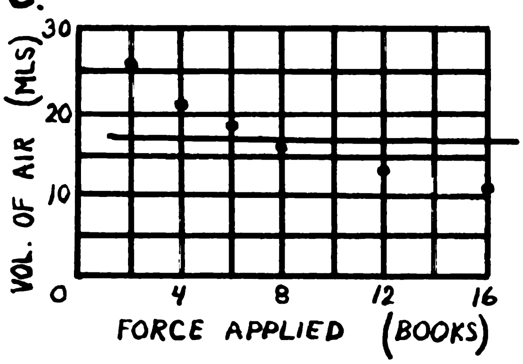
A.



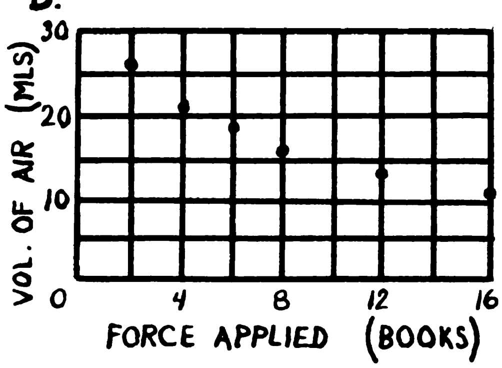
B.



C.



D.



78. In this experiment, which of the following were kept constant in the experiment?
    1. volume of air
    2. number of books
    3. the syringe
    4. none of the above

APPENDIX B

COMPLETE ITEM ANALYSIS DATA FOR THE ACCEPTED ITEMS ON THE PROCESS OF CONTROLLING VARIABLES

APPENDIX B

COMPLETE ITEM ANALYSIS DATA FOR THE ACCEPTED ITEMS ON THE PROCESS OF CONTROLLING VARIABLES

Item 25
The correct option is 4

ITEM RESPONSE PATTERN

	1	2	3	4*	5	Omit
Upper 27%	3 20%	2 13%	0 0%	8 53%	2 13%	0 0%
Middle 46%	6 23%	2 8%	1 4%	11 42%	3 12%	3 12%
Lower 27%	6 40%	2 13%	2 13%	2 13%	2 13%	1 7%
TOTAL	15 27%	6 11%	3 5%	21 38%	7 13%	4 7%

Index of difficulty .63

Index of discrimination .40

Item 27
The correct option is 4

ITEM RESPONSE PATTERN

		1	2	3	4*	Omit
Upper	27%	1 7%	1 7%	0 0%	13 87%	0 0%
Middle	46%	1 4%	3 12%	1 4%	19 73%	2 8%
Lower	27%	2 13%	3 20%	1 7%	8 53%	1 7%
TOTAL		4 7%	7 13%	2 4%	40 71%	3 5%

Index of difficulty .29

Index of discrimination .34

Item 29
The correct option is 3

ITEM RESPONSE PATTERN

		1	2	3*	4	Omit
Upper	27%	1 7%	5 33%	8 53%	1 7%	0 0%
Middle	46%	3 12%	4 15%	12 46%	4 15%	3 12%
Lower	27%	0 0%	8 53%	2 13%	3 20%	2 13%
TOTAL		4 7%	17 30%	22 39%	8 14%	5 9%

Index of item difficulty .60

Index of item discrimination .40

Item 30
The correct option is 4

ITEM RESPONSE PATTERN

		1	2	3	4*	Omit
Upper	27%	2 13%	2 13%	1 7%	10 67%	0 0%
Middle	46%	6 23%	3 12%	4 15%	11 42%	2 8%
Lower	27%	1 7%	1 7%	6 40%	5 33%	2 13%
TOTAL		9 16%	6 11%	11 20%	26 46%	4 7%

Index of item difficulty .54

Index of item idscrimination .34

Item 31
The correct option is 3

ITEM RESPONSE PATTERN

		1	2	3*	4	5	Omit
Upper	27%	3 20%	1 7%	8 53%	1 7%	2 13%	0 0%
Middle	46%	6 23%	1 4%	9 35%	3 12%	6 23%	1 4%
Lower	27%	5 33%	1 7%	1 7%	5 33%	0 0%	3 20%
TOTAL		14 25%	3 5%	18 32%	9 16%	8 14%	4 7%

Index of item difficulty .67

Index of item discrimination .46

Item 32
The correct option is 2

ITEM RESPONSE PATTERN

	1	2*	3	4	5	Omit
Upper 27%	0 0%	12 80%	1 7%	1 7%	1 7%	0 0%
Middle 46%	1 4%	18 69%	2 8%	3 12%	1 4%	1 4%
Lower 27%	1 7%	5 33%	1 7%	1 7%	3 20%	4 27%
TOTAL	2 4%	35 63%	4 7%	5 9%	5 9%	5 9%

Index of item difficulty .38

Index of item discrimination .47

Item 34
The correct option is 3

ITEM RESPONSE PATTERN

	1	2	3*	4	Omit
Upper 27%	1 7%	4 27%	10 67%	0 0%	0 0%
Middle 46%	3 12%	8 31%	10 38%	5 19%	0 0%
Lower 27%	2 13%	3 20%	3 20%	4 27%	3 20%
TOTAL	6 11%	15 27%	23 41%	9 16%	3 5%

Index of item difficulty .59

Index of item discrimination .47

Item 35
The correct option is 3

ITEM RESPONSE PATTERN

	1	2	3*	4	5	Omit
Upper 27%	2 13%	0 0%	13 87%	0 0%	0 0%	0 0%
Middle 46%	2 8%	0 0%	16 26%	4 15%	3 12%	1 4%
Lower 27%	4 27%	1 7%	7 47%	0 0%	0 0%	3 20%
TOTAL	8 14%	1 2%	36 64%	4 7%	3 5%	4 7%

Index of item difficulty .35

Index of item discrimination .40

Item 36
The correct option is 1

ITEM RESPONSE PATTERN

	1*	2	3	4	5	Omit
Upper 27%	10 67%	1 7%	1 7%	2 13%	1 7%	0 0%
Middle 46%	17 65%	0 0%	2 8%	6 23%	1 4%	0 0%
Lower 27%	4 27%	3 20%	1 7%	4 27%	0 0%	3 20%
TOTAL	31 55%	4 7%	4 7%	12 21%	2 4%	3 5%

Index of item difficulty .44

Index of item discrimination .40

Item 63
The correct option is 4

ITEM RESPONSE PATTERN

	1	2	3	4*	Omit
Upper 27%	0 0%	2 13%	1 7%	12 80%	0 0%
Middle 46%	5 19%	0 0%	5 19%	16 62%	0 0%
Lower 27%	2 13%	2 13%	4 27%	4 27%	3 20%
TOTAL	7 13%	4 7%	10 18%	23 57%	3 5%

Index of item difficulty .43

Index of item discrimination .53

Item 64
The correct option is 3

ITEM RESPONSE PATTERN

	1	2	3*	4	Omit
Upper 27%	1 7%	0 0%	14 93%	0 0%	0 0%
Middle 46%	5 19%	4 15%	16 62%	1 4%	0 0%
Lower 27%	3 20%	4 27%	4 27%	1 7%	3 20%
TOTAL	9 16%	8 14%	34 61%	2 4%	3 5%

Index of item difficulty .39

Index of item discrimination .66

Item 65
The correct option is 1

ITEM RESPONSE PATTERN

		1*	2	3	4	Omit
Upper	27%	11 73%	3 20%	1 7%	0 0%	0 0%
Middle	46%	15 58%	5 19%	4 15%	2 8%	0 0%
Lower	27%	4 27%	2 13%	4 27%	2 13%	3 20%
TOTAL		30 54%	10 18%	9 16%	4 7%	3 5%

Index of item difficulty .46

Index of item discrimination .46

Item 78
The correct option is 3

ITEM RESPONSE PATTERN

		1	2	3*	4	Omit
Upper	27%	4 27%	1 7%	9 60%	1 7%	0 0%
Middle	46%	6 23%	2 8%	15 58%	2 8%	1 4%
Lower	27%	3 20%	2 13%	4 27%	3 20%	3 20%
TOTAL		13 23%	5 9%	28 50%	6 11%	4 7%

Index of item difficulty .50

Index of item discrimination .33

APPENDIX C

COMPLETE ITEM ANALYSIS DATA FOR THE ACCEPTED ITEMS ON THE PROCESS OF INTERPRETING DATA

APPENDIX C

COMPLETE ITEM ANALYSIS DATA FOR THE ACCEPTED ITEMS ON THE PROCESS OF INTERPRETING DATA

Item 3
The correct option is 3

ITEM RESPONSE PATTERN

	1	2	3*	4	5	Omit
Upper 27%	0 0%	2 13%	11 73%	1 7%	0 0%	1 7%
Middle 46%	0 0%	5 19%	21 81%	0 0%	0 0%	0 0%
Lower 27%	1 7%	5 33%	7 47%	0 0%	0 0%	2 13%
TOTAL	1 2%	12 21%	39 70%	1 2%	0 0%	3 5%

Index of item difficulty .30

Index of item discrimination .26

Item 4
The correct option is 3

ITEM RESPONSE PATTERN

	1	2	3*	4	5	Omit
Upper 27%	1 7%	5 33%	5 33%	3 20%	0 0%	1 7%
Middle 46%	7 27%	3 12%	9 35%	7 27%	0 0%	0 0%
Lower 27%	5 33%	1 7%	2 13%	5 33%	0 0%	2 13%
TOTAL	13 23%	9 16%	16 29%	15 27%	0 0%	3 5%

Index of item difficulty .71

Index of item discrimination .20

Item 5
The correct option is 3

ITEM RESPONSE PATTERN

	1	2	3*	4	5	Omit
Upper 27%	0 0%	3 20%	6 40%	5 33%	0 0%	1 7%
Middle 46%	3 12%	4 15%	11 42%	7 27%	1 4%	0 0%
Lower 27%	3 20%	1 7%	3 20%	5 33%	1 7%	2 13%
TOTAL	6 11%	8 14%	20 36%	17 30%	2 4%	3 5%

Index of item difficulty .64

Index of item discrimination .20

Item 6
The correct option is 3

ITEM RESPONSE PATTERN

	1	2	3*	4	5	Omit
Upper 27%	0	0	11	3	0	1
	0%	0%	73%	20%	0%	7%
Middle 46%	0	3	16	7	0	0
	0%	12%	62%	27%	0%	0%
Lower 27%	3	0	5	4	0	3
	20%	0%	33%	27%	0%	20%
TOTAL	3	3	32	14	0	4
	5%	5%	57%	25%	0%	7%

Index of item difficulty .42

Index of item discrimination .40

Item 13
The correct option is 3

ITEM RESPONSE PATTERN

	1	2	3*	4	5	Omit
Upper 27%	0	2	11	1	0	1
	0%	13%	73%	7%	0%	7%
Middle 46%	2	2	16	3	1	2
	8%	8%	62%	12%	4%	8%
Lower 27%	2	1	2	4	3	3
	13%	7%	13%	27%	20%	20%
TOTAL	4	5	29	8	4	6
	7%	9%	52%	14%	7%	11%

Index of item difficulty .48

Index of item discrimination .60

Item 23
The correct option is 3

ITEM RESPONSE PATTERN

	1	2	3*	4	Omit
Upper 27%	2 13%	3 20%	5 33%	4 27%	1 7%
Middle 46%	6 23%	2 8%	12 46%	6 23%	0 0%
Lower 27%	7 47%	2 13%	1 7%	2 13%	3 20%
TOTAL	15 27%	7 13%	18 32%	12 21%	4 7%

Index of item difficulty .68

Index of item discrimination .26

Item 26
The correct option is 3

ITEM RESPONSE PATTERN

	1	2	3*	4	5	Omit
Upper 27%	0 0%	3 20%	5 33%	3 20%	2 13%	2 13%
Middle 46%	1 4%	11 42%	6 23%	4 15%	3 12%	1 4%
Lower 27%	1 7%	5 33%	0 0%	3 20%	3 20%	3 20%
TOTAL	2 4%	19 34%	11 20%	10 18%	8 14%	6 11%

Index of item difficulty .81

Index of item discrimination .33

Item 37
The correct option is 2

ITEM RESPONSE PATTERN

	1	2*	3	Omit
Upper 27%	4 27%	11 73%	0 0%	0 0%
Middle 46%	9 35%	10 38%	6 23%	1 4%
Lower 27%	8 53%	3 20%	2 13%	2 13%
TOTAL	21 38%	24 43%	8 14%	3 5%

Index of item difficulty .57

Index of item discrimination .53

Item 38
The correct option is 1

ITEM RESPONSE PATTERN

	1*	2	3	Omit
Upper 27%	14 93%	1 7%	0 0%	0 0%
Middle 46%	16 62%	7 27%	2 8%	1 4%
Lower 27%	9 60%	3 20%	1 7%	2 13%
TOTAL	39 70%	11 20%	3 5%	3 5%

Index of item difficulty .30

Index of item discrimination .33

Item 39
The correct option is 2

ITEM RESPONSE PATTERN

	1	2*	3	Omit
Upper 27%	3 20%	10 67%	2 13%	0 0%
Middle 46%	4 15%	19 73%	2 8%	1 4%
Lower 27%	4 27%	7 47%	2 13%	2 13%
TOTAL	11 20%	36 64%	6 11%	3 5%

Index of item difficulty .36

Index of item discrimination .20

Item 40
The correct option is 3

ITEM RESPONSE PATTERN

	1	2	3*	Omit
Upper 27%	7 47%	0 0%	8 53%	0 0%
Middle 46%	14 54%	2 8%	9 35%	1 4%
Lower 27%	5 33%	3 20%	5 33%	2 13%
TOTAL	26 46%	5 9%	22 39%	3 5%

Index of item difficulty .60

Index of item discrimination .20

Item 45
The correct option is 2

ITEM RESPONSE PATTERN

	1	2*	3	4	Omit
Upper 27%	4 27%	11 73%	0 0%	0 0%	0 0%
Middle 46%	6 23%	16 62%	2 8%	1 4%	1 4%
Lower 27%	4 27%	6 40%	1 7%	2 13%	2 13%
TOTAL	14 25%	33 59%	3 5%	3 5%	3 5%

Index of item difficulty .40

Index of item discrimination .33

Item 46
The correct option is 2

ITEM RESPONSE PATTERN

	1	2*	3	4	Omit
Upper 27%	0 0%	13 87%	1 7%	1 7%	0 0%
Middle 46%	0 0%	19 73%	4 15%	2 8%	1 4%
Lower 27%	0 0%	4 27%	4 27%	5 33%	2 13%
TOTAL	0 0%	36 64%	9 16%	8 14%	3 5%

Index of item difficulty .35

Index of item discrimination .60

Item 47
The correct option is 3

ITEM RESPONSE PATTERN

	1	2	3*	4	Omit
Upper 27%	0 0%	0 0%	14 93%	1 7%	0 0%
Middle 46%	3 12%	1 4%	19 73%	2 8%	1 4%
Lower 27%	1 7%	1 7%	7 47%	4 27%	2 13%
TOTAL	4 7%	2 4%	40 71%	7 13%	3 5%

Index of item difficulty .29

Index of item discrimination .46

Item 48
The correct option is 1

ITEM RESPONSE PATTERN

	1*	2	3	4	Omit
Upper 27%	10 67%	4 27%	0 0%	1 7%	0 0%
Middle 46%	9 35%	11 42%	4 15%	1 4%	1 4%
Lower 27%	1 7%	5 33%	6 40%	1 7%	2 13%
TOTAL	20 36%	20 36%	10 18%	3 5%	3 5%

Index of item difficulty .64

Index of item discrimination .60

Item 49
The correct option is 3

ITEM RESPONSE PATTERN

	1	2	3*	4	Omit
Upper 27%	4 27%	1 7%	7 47%	2 13%	1 7%
Middle 46%	5 19%	2 8%	13 50%	5 19%	1 4%
Lower 27%	1 7%	3 20%	4 27%	5 33%	2 13%
TOTAL	10 18%	6 11%	24 43%	12 21%	4 7%

Index of item difficulty .57

Index of item discrimination .20

Item 66
The correct option is 4

ITEM RESPONSE PATTERN

	1	2	3	4*	Omit
Upper 27%	1 7%	3 20%	2 13%	9 60%	0 0%
Middle 46%	3 12%	4 15%	3 12%	15 58%	1 4%
Lower 27%	1 7%	3 20%	4 27%	5 33%	2 13%
TOTAL	5 9%	10 18%	9 16%	29 52%	3 5%

Index of item difficulty .48

Index of item discrimination .27

Item 67
The correct option is 1

ITEM RESPONSE PATTERN

	1*	2	3	4	Omit
Upper 27%	11 73%	3 20%	1 7%	0 0%	0 0%
Middle 46%	17 65%	0 0%	6 23%	2 8%	1 4%
Lower 27%	7 47%	4 27%	1 7%	1 7%	2 13%
TOTAL	35 63%	7 13%	8 14%	3 5%	3 5%

Index of item difficulty .37

Index of item discrimination .26

Item 68
The correct option is 2

ITEM RESPONSE PATTERN

	1	2*	3	4	Omit
Upper 27%	0 0%	14 93%	1 7%	0 0%	0 0%
Middle 46%	3 12%	17 65%	4 15%	1 4%	1 4%
Lower 27%	0 0%	5 33%	6 40%	2 13%	2 13%
TOTAL	3 5%	36 64%	11 20%	3 5%	3 5%

Index of item difficulty .35

Index of item discrimination .60

Item 73
The correct option is 4

ITEM RESPONSE PATTERN

		1	2	3	4*	Omit
Upper	27%	0 0%	0 0%	0 0%	15 100%	0 0%
Middle	46%	1 4%	0 0%	2 8%	22 85%	1 4%
Lower	27%	3 20%	2 13%	1 7%	7 47%	2 13%
TOTAL		4 7%	2 4%	3 5%	44 79%	3 5%

Index of item difficulty .21

Index of item discrimination .53

Item 74
The correct option is 3

ITEM RESPONSE PATTERN

		1	2	3*	4	Omit
Upper	27%	0 0%	0 0%	15 100%	0 0%	0 0%
Middle	46%	0 0%	1 4%	24 92%	0 0%	1 4%
Lower	27%	2 13%	3 20%	8 53%	0 0%	2 13%
TOTAL		2 4%	4 7%	47 84%	0 0%	3 5%

Index of item difficulty .16

Index of item discrimination .47

Item 75
The correct option is 4

ITEM RESPONSE PATTERN

		1	2	3	4	5	Omit
Upper	27%	0	0	0	13	1	1
		0%	0%	0%	87%	7%	7%
Middle	46%	1	1	4	16	3	1
		4%	4%	15%	62%	12%	4%
Lower	27%	0	0	2	7	4	2
		0%	0%	13%	47%	27%	13%
TOTAL		1	1	6	36	8	4
		2%	2%	11%	64%	14%	7%

Index of item difficulty .36

Index of item discrimination .40