

This is to certify that the
thesis entitled

DEVELOPMENT OF THE SCIENCE PROCESSES
TEST (TSPT)

presented by

Robert R. Ludeman

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in EDUCATION

Edward L. Smith

Major professor

Date November 8, 1974

ABSTRACT

DEVELOPMENT OF THE SCIENCE PROCESSES TEST TSPT

by

Robert R. Ludeman

PROBLEM

There is considerable evidence that the technology of educational evaluation has not kept pace with developments in other areas of the educational enterprise. This study involved the development of a test of science processes using a method of item selection which replaced the customary panel of judges who pass on the items' validity with an objective method of item selection based on an external criterion. Some of the characteristics of the resulting test are examined.

LITERATURE

The literature is examined with reference to several issues relevant to the mechanics of test construction, including speeded vs power tests, the test blueprint, the optimum number of alternatives, item order, acceptable difficulty level, and item discrimination. A short survey of recently developed science process tests is presented with the method of validation used in each case. The concern expressed by some testing authorities over traditional methods of validation is examined and the external criterion referenced method developed by Fyffe and Robison and used in this study is reviewed.

PROCEDURE

The item improvement phase of the study involved addition to and revision of the items developed by Fyffe and Robison using the item analysis data generated by their study. Two additional item tryout and revision cycles were required before item analysis indicated the item pool of 61 items to be of adequate quality. The result was known as The Science Processes Test (TSPT) form C.

The validation phase of the study consisted of the administration of three tests to the validation sample which was composed of 52 sixth grade students. The three tests were a subset of the Individual Competency Measures taken from the Science - A Process Approach elementary science program, TSPT form C, and the Science Research Associates (SRA) test. The correlation of students' scores on each of the form C items with their scores on the four subtests of the Individual Competency Measures was computed and hypothesis one was tested. Hypothesis one was that scores on each item of form C would exhibit a significantly higher correlation with scores on one of the subtests of the Individual Competency Measures than with any other subtest.

The scores on the Individual Competency Measures served as the external criterion measure for selecting the upper and lower 27 percent groups needed to calculate the item discrimination indices. Form C items were selected to be included in form D based on the requirement that this external criterion referenced discrimination index have a minimum value of 0.20. Thirty-six items from form C which met this requirement were included in TSPT form D.

The correlation of form D scores with the Individual Competency Measures scores was computed. Hypothesis two, that form D scores were

more highly correlated with the Individual Competency Measures scores than with the SRA Science test scores was tested.

Norming data for TSPT form D was obtained by administering it to a random sample of 1301 sixth grade students. The preparation of a test manual for TSPT form D completed this study.

RESULTS

The correlation of TSPT form D scores with the Individual Competency Measures scores is 0.83, which is significant well beyond the 0.001 level, and demonstrates that the external criterion referenced method of test development used is a fruitful approach to test construction. The hypothesis that items could be objectively assigned to the Individual Competency Measures subtests was not supported and the intercorrelations among the Individual Competency Measures subtests cast such doubt on their independence that no further reference was made to the supposed subscales. The hypothesis that TSPT scores would be more highly correlated with the Individual Competency Measures scores than with the SRA Science test scores was also not supported. The high correlation between the Individual Competency Measures and the SRA Science test scores raises the question whether process tests, which the former is claimed to be, and factual knowledge tests, which the latter is claimed to be, do indeed lead to greatly different results.

CONCLUSIONS

Although the value of TSPT will only become apparent as it is used, its high correlation with the Individual Competency Measures and its quality as indicated by the test statistics suggest that it should be of value to those concerned with science process evaluation and that

the method of test construction used may be of value to those concerned with test construction.

RECOMMENDATIONS

In view of the results of this study it is recommended that:

1. TSPT be used and evaluated by researchers.
2. Further use be made of the objective method of test development used in this study.
3. Additional research be done to test the independence of process test subscales.
4. Additional research be done to distinguish between process ability and factual knowledge.

DEVELOPMENT OF THE
SCIENCE PROCESSES TEST (TSPT)

by

Robert R. Ludeman

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

College of Education

1974

2/10/74

© Copyright by
ROBERT R. LUDEMAN
1974

ACKNOWLEDGEMENTS

Many people have contributed toward the success of this study. Over 1500 school children and some 100 teachers and administrators have helped in the study. Of these, special mention must be made of the students in the Pierce Community School who made up the validation sample and their teachers who did not complain at my extensive disruptions.

Grateful appreciation is also due to Drs. Sherwood K. Haynes, Robert L. Ebel, and Glenn D. Berkheimer, members of my guidance committee. Each member contributed from his expertise, insights and advice which aided my personal growth and added to the value of this study.

A special note of thanks goes to the chairmen of my guidance committee, Dr. Richard J. McLeod who suggested the study and gave valuable guidance early in the work and to Dr. Edward L. Smith whose interest and willingness to go beyond the call of duty in giving advice and guidance contributed significantly to the quality of the study.

Special thanks is also due to the administration of Andrews University without whose encouragement and financial support the work would never have been undertaken.

Most of all the greatest thanks goes to my dear wife who endured it all for me and without whose love it would have been neither possible nor desirable.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	ii
LIST OF TABLES	v
LIST OF APPENDICES	vi
CHAPTER	
I. THE PROBLEM	1
Background	1
The Need for the Study	3
The Purpose of the Study	3
Initial Considerations	4
Hypotheses to be Tested	5
Test Instruments Used	7
Assumptions	8
Limitations	10
Overview of the Thesis	11
Footnotes	12
II. REVIEW OF THE LITERATURE	15
Background	15
Process Evaluation	16
Test Construction	16
Other Process Tests	19
The Need for External Criterion Referenced Validation	22

CHAPTER	Page
The Work of Fyffe and Robison	23
Summary	24
Footnotes	26
III. PROCEDURE	31
Item Improvement	31
Validation Phase	35
Multiple Regression Analysis	42
Norming TSPT Form D	43
Test Manual Preparation	45
Summary	45
Footnotes	48
IV. ANALYSIS OF RESULTS	49
Item Improvement	49
Validation	54
Norming TSPT Form D	65
Summary	68
Footnotes	69
V. SUMMARY AND CONCLUSIONS	70
Summary	70
Conclusions	72
Implications for Further Research	74
Footnotes	77
BIBLIOGRAPHY	78
APPENDICES	83

LIST OF TABLES

Table	Page
1. Item Subtest Assignments	50
2. TSPT Test Statistics	51
3. TSPT Form A Correlation Table	52
4. TSPT Form C Subtest Correlations	54
5. TSPT Form C Item Assignments	56
6. TSPT Form D Item Selection Criteria	57
7. SRA Test Statistics	58
8. TSPT, ICM - TSPT, SRAS Correlation Comparison	60
9. Individual Competency Measures Subtest Intercorrelations	61
10. Multiple Regression Analysis	63
11. Norming Schools Characteristics	65
12. TSPT Form D Test Statistics	66
13. Norming Sample Frequency Distribution	67

LIST OF APPENDICES

Appendix	Page
I-A. One Individual Competency Measure from SAPA	83
I-B. Integrated Processes of SAPA	85
I-C. Listing of Individual Competency Measures	87
III-A. The Minimum Level of Discrimination - Conventional Item Analysis	89
III-B. Directions for Administering TSPT Form D	90
III-C. TSPT Form D Test Manual	93
IV-A. TSPT Form A	107
IV-B. Form A Subtest Assignments	132
IV-C. Item Analysis Form A	133
IV-D. TSPT Form C	146
IV-E. Item Analysis Form C	166
IV-F. Validation Sample Scores	175
IV-G. TSPT Form C Items - Individual Competency Measures Subtest Correlations	177
IV-H. TSPT Form D	180
IV-I. Norming Area Map	194
IV-J. Item Analysis Form C	196

CHAPTER I

THE PROBLEM

BACKGROUND

The educational theorists have long recognized the need to teach more than just factual knowledge in the schools but it was during the post-sputnik era that educational practice began to make significant progress in that direction.¹ It was at that time that the "acronym curriculum" of innovative science programs began to emerge. All of them to a greater or lesser degree claim to teach the "higher mental processes."² Science - A Process Approach (SAPA), the elementary school science program developed by the American Association for the Advancement of Science, has been one of the leaders, for the whole of SAPA is built around the processes which the developers have identified as basic to science.³

As the new courses came into common use it became apparent that there were few, if any, tests available for assessing knowledge of these "higher mental processes." It also became painfully apparent that good process test items are not easy to write.^{4,5} Some spokesmen were moved to call for new methods of evaluation^{6,7,8} and new research relative to assessment of achievement at the higher cognitive levels.^{9,10,11}

Anticipating the need for evaluation, the designers of SAPA constructed the Individual Competency Measures. It was originally intended that the Individual Competency Measures would be administered to one student at a time. The teacher would verbally set the task, which occasionally involved the use of physical objects, and then he would

observe the student's behavior and record his competence in using the process skills required to perform the task. Student performance on each required task is described in detail for the teacher so that he can judge the student's performance.¹² Appendix I-A contains a sample of the Individual Competence Measures. Since the student is asked to demonstrate his knowledge of the processes in much the same setting as that in which they were learned, it is reasonable to ascribe to them "primary" or "direct" validity.¹³ The difficulty with the Individual Competency Measures is that since they are administered individually they are too time consuming to be used very widely. Although Group Competency Measures have been developed for administration to from three to six students at a time, this is a compromise which does not solve the problem of time efficiency.

Even though the Individual Competency Measures are too time consuming to be used extensively, because of their validity Fyffe¹⁴ and Robison¹⁵ recognized their potential as a standard against which to compare a more time efficient test of science processes. They took the initial steps in the development of such a test by generating a pool of test items each of which was validated by correlation with the Individual Competency Measures relating to one or more of the following four Integrated Processes as defined by SAPA: Interpreting Data, Controlling Variables, Formulating Hypotheses, and Defining Operationally. These processes are defined in Appendix I-B of this study. Since their work was not intended to produce a usable test, there is no way of evaluating the merit of Fyffe and Robison's approach. This study will develop such a test and a tentative evaluation of the approach will be attempted.

THE NEED FOR THE STUDY

A number of individuals have recognized the need for time efficient tests of children's ability to use the science processes^{16,17} and high quality, time efficient pencil and paper tests have been prepared which claim to assess ability to use these processes.^{18,19} However, as is true for so many other tests, they almost always base their claim of validity solely on "expert" opinion. This is considered to be a serious weakness.^{20,21,22} One reason for this practice is that if another measure is to be used as an external standard for validation, Ebel points out that "...it should always exemplify a measurement procedure clearly superior to (i.e., more relevant, more precise than) that embodied in the test in question."²³ Obviously, in most cases, if such a measure exists, it will be used and there is no need to construct another. However, in this case, due to their time inefficiency, the Individual Competency Measures are not practical to use directly but, by the nature of their construction and format, they do meet Ebel's requirements for superior relevance and precision. Thus, in testing ability to use the science processes the opportunity does exist for the construction of a test which does not depend solely on "expert opinion" for its validation. This opportunity is pursued in this study.

THE PURPOSE OF THE STUDY

The purpose of this study is to develop a test, The Science Processes Test (TSPT), using item selection based on item discrimination referenced to an external criterion, and to evaluate the test's performance. This method is described in detail later in this chapter. The external criterion that will be used for determining the item discrimination is the Individual Competency Measures of SAPA. TSPT is intended to

be a research instrument of sufficient quality to be usable by researchers in science education for assessing students' ability to use the integrated processes of Interpreting Data, Controlling Variables, Formulating Hypotheses and Defining Operationally as defined by SAPA. The manual for TSPT form D has been prepared in accordance with the American Educational Research Association recommendations for such manuals.²⁴

Since, as has been previously mentioned, others have found it difficult to construct items which assess ability to use the processes of science, as additional assurance of test validity, students' performance on TSPT will be compared with their performance on the Individual Competency Measures and on the Science Research Associates (SRA) Science test, a test, which it is claimed, measures mainly factual knowledge.²⁵ If the TSPT scores are more closely correlated to the Individual Competency scores than to the SRA Science scores this will be taken as evidence that TSPT is more a test of science processes than of factual knowledge.

INITIAL CONSIDERATIONS

Early in the development certain decisions were made with reference to the development and final form of TSPT.

1. The time span required to administer the test would be no more than approximately 45 minutes.

2. The test would be of pencil and paper multiple choice format. The reason for decisions 1 and 2 is the requirement that the test be easily administered without the requirement of a special testing period and without special facilities, equipment, or training of the test administrator.

3. The test would be a non-paced power test not having any time limitation. There is evidence that timing this type of test is not wise.

Both decisions 2 and 3 above required that items built around projected pictures in the original item pool had to be rewritten. In some instances, in order to minimize the reading required, printed pictures were substituted.

4. The major portion of the study would not be attempted until the items in the item pool appeared to be of adequate technical quality to be useful as test items. The criteria for making this judgment are presented in Chapter III.

5. The subjects used would be limited to only one grade level. The reason for this decision is the elimination of as many variables as possible. The sixth grade was chosen because typically it is the last grade in which the SAPA materials are used and the integrated processes are given increased emphasis in the later grades.

HYPOTHESES TO BE TESTED

Although the major portion of this study is concerned with the development of TSPT, the section on validation does have an experimental aspect with the following hypotheses to be tested:

1. The Integrated Process which a given test item assesses will be indicated by the students' scores on the item having a significantly higher correlation with their scores on that Integrated Process subtest than on any other subtest of the Individual Competency Measures.

2. Student scores on TSPT will have a significantly higher correlation with their scores on the Individual Competency Measures than the correlation they have with the SRA Science test.

THE EXTERNAL CRITERION REFERENCED METHOD OF TEST DEVELOPMENT

As used in this study, this method of test development deviates from the typical method of test improvement through item analysis in two important respects:

1. The "upper 27 percent" and the "lower 27 percent" groups used in the item analysis are determined with reference to the external criterion scores as opposed to the conventional procedure which uses the scores on the test under development to determine these groups. This procedure suggested by Fyffe²⁶ and Robison²⁷ provides assurance that items will be selected on which students who know the material assessed by the criterion test do well and students who do not know the material assessed by the criterion test do poorly. In other words, it provides assurance that the item discriminates on the basis of the external criterion. If one has this assurance, it is expected that students' performance on a test composed of such items will correlate highly with their performance on the external criterion. The minimum value used for this discrimination in this study was 0.2.

2. In order to have further assurance of a high correlation with the external criterion, a further requirement is used in this study. It is that student scores on the item will have a minimum correlation with their criterion test scores of 0.2. In most cases this latter requirement is not necessary. If the discrimination requirement is met, the correlation requirement will be met.

The reason the discrimination and correlation requirements are lower than that usually used is that for the usual method of item analysis, the item under consideration has contributed to the total score and so the value is artificially inflated. This is not true when the external criterion is used. Table 6 contains empirical evidence that, at least for this study, 0.2 is an appropriate value.

TEST INSTRUMENTS USED

The Individual Competency Measures

A set of tests designed by SAPA to be administered to one student at a time.²⁸ The testor verbally sets a task which frequently involves a hands-on manipulation of a physical object such as the use of a stop-watch, a balance, or a meter stick to make measurements. The testor observes the student's behavior and records his competence in using the process skills to perform the task he has set. Acceptable student performance on each task is described in detail for the testor so that he can quantitatively rate the student's performance. A typical sample of the Individual Competency Measures is included in Appendix I-A. A listing of the Individual Competency Measures considered for use in this study is included in Appendix I-C.

SRA Science Test

Science Research Associates Achievement Series: Science (blue version) form D.²⁹ This test was chosen because it is of high quality and, most important, it is criticized as follows by one reviewer.³⁰

"The test appears to measure primarily a mastery of science content. It focuses mainly upon knowledge and to a more limited extent upon understanding. It is not appreciably concerned with processes of science or with the problem centered approach."

It is the lack of concern for the processes which makes the test ideal for this study for this means it should be measuring something distinctly different from what the Individual Competency Measures measure. This "something" for the purposes of this study will be referred to as "factual knowledge."

SRA Reading Test

Science Research Associates Achievement Series: Reading (blue version) form D.³¹ This test was chosen because it was the companion test for the SRA Science test. In an effort to shorten the total test, there were some items which SRA scored as both science and reading items. These items were dropped from the reading test in order that the reading test would measure as much "non-science" as possible.

Fry Readability Formula

Due to the fact that reading scales are typically intended for use with textual material, it was felt that in this case, use of one of the more complex reading scales was not warranted. The Fry Readability Formula is an easy to use readability formula based on grammatical complexity and vocabulary.³² The rule followed was that only the correct alternative was considered in the calculation and numbers were considered to contain one syllable for each digit. Fry places the uncertainty of grade level determination for his scale at approximately one grade level. The uncertainty is probably higher in this situation, but at least some indication is given of the probable reading level.

ASSUMPTIONS

Probably the most important assumption of the study is that the Individual Competency Measures are valid measures of students' ability to use the science processes. Since correlation with performance on the Individual Competency Measures is used as the criterion for judging whether or not an item should be included in TSPT and also as the criterion for judging to what extent TSPT measures students' ability to use the processes of science, this assumption underlies the entire study.

Second, it is assumed that the inflation in correlation of TSPT with the Individual Competency Measures which is bound to result from the above procedure of using the same data for construction of the test and for analysis of the test will not be serious enough to alter the outcome of the study. This assumption will be discussed further in Chapter V.

Third, it is assumed that the SRA Science test does not measure ability to use the processes of science.

Fourth, it is assumed that the validation sample will contain both students who possess the ability to use the science processes and students who possess factual knowledge with respect to science but that these abilities are possessed quite independently of one another.

Fifth, it is assumed that for the validation phase of the study individual student's ability to use the science processes did not change significantly between the time when it was measured using the Individual Competency Measures and the time of administration of TSPT form C. This time interval could not be reduced below approximately one and one-half months due to the time required to administer the Individual Competency Measures.

Sixth, it is assumed that the learning effect or carry-over from one test to another will not significantly affect students' performance on the tests.

Seventh, it is assumed that the vocabulary and context of the test is sufficiently general that its usefulness will not be limited to students who have studied the SAPA materials.

Eighth, it is assumed that the use of standard statistical procedures and item analysis procedures are applicable to this situation.

LIMITATIONS

This study is limited to the extent that the preceding assumptions are invalid. Further, this study is limited in its interpretation of what really constitutes the science processes to the interpretation used in SAPA. But the vocabulary used in TSPT is not unique to SAPA so its use will not be necessarily limited to students familiar with SAPA materials. Similarly, this study is limited in its interpretation of what really constitutes factual knowledge to the abilities assessed by the SRA Science test.

The correlation of TSPT scores with the Individual Competency Measures scores is expected to be high and the correlation of TSPT scores with the SRA Science test scores is expected to be low. Since this difference will be taken as evidence that TSPT measures ability to use the science processes, to the extent that the Individual Competency Measures measure factual knowledge, and to the extent that the SRA Science test measures the science processes, the correlation of the Individual Competency Measures with the SRA Science test will be high and contamination will be introduced. Similarly, if the validation sample contains students who do not possess the ability to use the science processes, or if they do not possess factual knowledge, or if these abilities do not exist independently of each other, again the correlation of the Individual Competency Measures with the SRA Science test will be inflated. Either of these effects will tend to obscure the expected results, that TSPT measures science processes to a greater extent than it measures factual knowledge.

OVERVIEW OF THE THESIS

In this chapter the following topics have been presented: The background including the work which has lead up to this study, the need, the purpose intended to be accomplished, the hypotheses to be addressed, a brief description of the test instruments to be used, the assumptions on which the study is based, and finally, the limitations of the study.

In Chapter II a review of pertinent literature will be presented, including a brief review of the trend toward process education, the SAPA program, the effect the new programs have had on evaluation, and attempts to improve evaluative techniques.

Chapter III describes the procedure used to conduct this study. The first step is the item tryout and improvement. The validation part of the study introduces the unique method of test development used which is given the descriptive title of the external criterion referenced validation method of test development. Finally the norming procedure and the test manual preparation are described.

Chapter IV presents the analysis of the data obtained. The item analysis data from the tryouts is presented first. In connection with the validation study the statistical hypotheses are tested and the data reduction involved in the development of TSPT form D is presented. Finally, the norming data for publication in the test manual is presented.

Chapter V contains a summary of the findings, the conclusions arrived at, and a discussion of the implications of the study.

FOOTNOTES

¹Henry P. Cole, "Process Curricula and Creativity Development," Journal of Creative Behavior, 3, 253 (Fall 1969).

²Terry Borton, "What's Left When School's Forgotten?" Saturday Review, 53, 69-71, 79 (April 18, 1970).

³American Association for the Advancement of Science, The Psychological Bases of Science - A Process Approach, AAAS Misc. Publication (1965).

⁴L. Lisonbee, "Testing, What For?" Science Teacher, 33, 27-29 (May 1966).

⁵Hulda Grobman, Evaluation Activities of Curriculum Projects, AERA Monograph Series on Curriculum Evaluation, No. 2 (Rand McNally, Chicago, Illinois, 1968).

⁶Max D. Engelhart and John M. Beck, "The Improvement of Tests," The 62nd Yearbook of the National Society for the Study of Education (Chicago: University of Chicago Press, 1963).

⁷Robert E. Stake and T. Denny, "Needed Concepts and Techniques for Utilizing More Fully the Potential of Evaluation," The 68th Yearbook of the National Society for the Study of Education, 2 (Chicago: University of Chicago Press, 1969).

⁸Richard B. Smith, "Approach to Measurement in the New Science Curriculum," Science Education, 53, 411-415 (December 1969).

⁹Bernard W. Benson and L. L. Young, "Development and Implementation of an Instrument to Assess Cognitive Performance in High School Biology; Assessment of Cognitive Transfer in Science Inventory," Journal of Research in Science Teaching, 8, 211-224 (1971).

¹⁰Robert H. Ennis, "Needed: Research in Critical Thinking," Educational Leadership, 21, 17-20 (1971).

¹¹Ralph W. Tyler, "Educational Evaluation: New Roles, New Means," The 68th Yearbook of the National Society for the Study of Education, 2, (Chicago: University of Chicago Press, 1969).

¹²American Association for the Advancement of Science, loc. cit.

¹³Robert L. Ebel, Essentials of Educational Measurement, (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1972), p. 438.

¹⁴Darrell W. Fyffe, The Development of Test Items for the Integrated Science Processes: Formulating Hypotheses and Defining Operationally, Unpublished Doctoral Dissertation, Michigan State University (1971).

¹⁵Richard Wayne Robison, The Development of Items which Assess the Processes of Controlling Variables and Interpreting Data, Unpublished Doctoral Dissertation, Michigan State University (1973).

¹⁶H. Grobman, "Curriculum Development and Evaluation," Journal of Educational Research, 64, 436-422 (July 1971).

¹⁷Ralph W. Tyler, "Resources, Models, and Theory in the Improvement of Research in Science Education," Journal of Research in Science Teaching, 5, 43 (1967).

¹⁸Jacqueline V. Mallison, "Review - Stanford Achievement Test: Science," in Oscar K. Buros, Seventh Mental Measurements Yearbook, 2 (Gryphon Press, Highland Park, New Jersey, 1972).

¹⁹Irvin J. Lehmann, "Review - Test of Academic Progress: Science," in Oscar K. Buros, Seventh Mental Measurements Yearbook, 2 (Gryphon Press, Highland Park, New Jersey, 1972), pp. 1243-45.

²⁰Warren G. Findley, "Purposes of School Testing Programs and Their Efficient Development," The 62nd Yearbook of the National Society for the Study of Education (Chicago: University of Chicago Press, 1963), p. 8.

²¹James R. Barclay, Controversial Issues in Testing (Boston: Houghton Mifflin, 1968), p. 60.

²²Robert E. Stake, "The Countenance of Educational Evaluation," Teachers College Record, 68, 523-540 (April 1967).

²³Robert L. Ebel, "Must all Tests be Valid?" American Psychologist, 16, 640-647 (October 1961).

²⁴American Educational Research Association Committee on Test Standards, Technical Recommendations for Achievement Tests (National Education Association, Washington, D.C., 1955).

²⁵Clarence H. Nelson, "Review - Science Research Associates Achievement Series: blue version," in Oscar K. Buros, Seventh Mental Measurements Yearbook, 2 (Gryphon Press, Highland Park, New Jersey, 1972), pp. 1231-33.

²⁶Fyffe, op. cit., p. 40.

²⁷Robison, op. cit., p. 51.

²⁸American Association for the Advancement of Science, loc. cit.

²⁹Science Research Associates Achievement Series: blue version, form D (Chicago: Science Research Associates, Inc.).

³⁰Nelson, loc. cit.

³¹Science Research Associates Achievement Series, loc. cit.

³²Edward B. Fry, Reading Instruction for Classroom and Clinic
(New York: McGraw Hill, 1972), p. 205.

CHAPTER II

REVIEW OF THE LITERATURE

The literature on recent developments in both teaching and testing is extensive and no attempt will be made to report exhaustively on either. Rather, in this chapter, the recent emphasis on science processes and the implications this emphasis has for testing will be documented. Also some other recent attempts to construct tests which assess students' ability to use the science processes will be presented.

BACKGROUND

Educators have long recognized the value of teaching students the procedures and strategies of inquiry used by scientists^{1,2} and as a result science process teaching has become the focal point for several curriculum projects³ with SAPA being one of the leaders.⁴ SAPA has identified eight basic processes and five integrated processes about which they have built their program. The basic processes are: observing, using space/time relationships, classifying, using numbers, measuring, communicating, predicting, and inferring. The integrated processes are: Interpreting Data, Controlling Variables, Formulating Hypotheses, Defining Operationally, and Experimenting.⁵ It should be emphasized that the integrated processes are claimed to include the basic processes and that the integrated process of experimenting is claimed to encompass all the other processes.⁶ Thus the first four integrated processes are of concern for this study.

PROCESS EVALUATION

It is generally conceded that "To teach without testing is unthinkable,"⁷ that objectives should be testable, and that evaluation should extend to all the outcomes to which the school addresses itself.⁸ And yet, many science educators assert that test development has not kept pace with the curriculum changes of the past decade.⁹⁻¹⁵ The result is that too often we are teaching what we are not testing, and testing what we are not teaching. Lisonbee¹⁶ points out that one probable reason for this situation is that the objectives as listed by many of the curriculum designers are often not testable. An obvious reason suggested by Grobman¹⁷ is the difficulty and expense of developing a good test. It has also been suggested that new approaches to testing are needed if process abilities are to be assessed.¹⁸⁻²¹

TEST CONSTRUCTION

Although the multiple choice test has had its detractors,²² it has emerged as the testing format of choice for most testing situations and will be the only format considered here. A rather standard methodology has evolved for test development and use and there are a number of good sources²³⁻²⁶ which describe the techniques in detail.

Speeding

The question whether a test should be a "speeded test" or a "power test" has been examined by a number of investigators.^{27,28} The consensus seems to be, "In some situations speed tests may be appropriate and valuable, but these situations seem to be the exception not the rule."²⁹ Those exceptions would be when time is a factor in the evaluation such as a typist's speed test. Otherwise, especially in situations where

careful thinking was involved, speeding has been found to reduce test reliability.³⁰ It was decided that TSPT should be a power test.

The Blueprint

Travers³¹ has suggested that in order to aid in achieving the desired balance among item types and concepts used, a two dimensional matrix or "blueprint" should be employed for assigning items to the test under construction. Others have suggested that the multidimensional matrix may be too awkward and time consuming and that perhaps a vector or one dimensional matrix composed of categories may be of more practical utility to the test constructor.^{32,33} For construction of TSPT, the four integrated processes of Interpreting Data, Controlling Variables, Formulating Hypotheses, and Defining Operationally were used as the test item categories.

Number of Alternatives

Tversky³⁴ has developed a mathematical proof based on certain assumptions relative to test characteristics and sample properties which indicates that use of three alternatives will maximize the discrimination of a multiple choice test. Costin³⁵ has submitted empirical evidence which indicates both discrimination and reliability show a slight increase when three alternatives are used as opposed to four alternatives. Opposing these findings, Ebel³⁶ has developed a formula based on different assumptions relative to the characteristics of the test and the sample which indicates that the maximum possible reliability for a 100 item test could continue to increase as the number of alternatives is increased, though the rate of increase drops rapidly beyond about four alternatives.

In view of the above findings, it is probably safe to say that the practical difference between using three or four alternatives would be small. It was therefore concluded that due to its wider acceptance, the four alternative format would be used.

Item Arrangement

Many investigators have examined the effect of item arrangement on test performance. Flaughner³⁷ has reported a statistically significant improvement in scores on a verbal test favoring the easy to hard arrangement although he questions the practical significance of this finding, and reports no effect on a mathematics test. The findings of Munz³⁸ are quite similar. More important, Brenner³⁹ reports item order did not significantly affect test reliability, difficulty, or discrimination. Marso and Klosner^{40,41} support these findings. Thus item difficulty was not used to decide on item order for TSPT.

Item Difficulty

There has apparently been a rather noticeable shift in thinking over the years among testing experts in relation to item difficulty. Symonds⁴² reported that maximum reliability would be achieved if the difficulty was close to 0.50. Davis⁴³ in a review of both theory and research agreed with this view. Adams⁴⁴ showed that the highest test reliability was achieved with items of middle difficulty levels. Wofford⁴⁵ reported that contrary to theoretical prediction, wider difficulty ranges (0.25 to 0.75) does not decrease reliability. Recently Davis⁴⁶ has indicated that a difficulty level near 0.5 is not as important as has been thought.

Item Discrimination

Kelly⁴⁷ originally proposed the use of the "upper 27 percent" and the "lower 27 percent" as the extreme groups for item analysis. Both Feld⁴⁸ and Wofford⁴⁹ later supported Kelly, although Wofford indicated that there was no difference in result when the total sample was used rather than the upper and lower 27 percent. Engelhart⁵⁰ compared a number of different indices which have been proposed for use as indicators of the ability of an item to discriminate. He concluded that the "D" index (the difference between the upper 27 percent and the lower 27 percent) was about as effective as any of the correlation type indices which have been proposed in identifying poor test items. It also has the advantage of being more indicative of the actual number of discriminations made. Thus Kelly's index seems to have stood the test of time and even the onslaught of hard to compute indices made usable by modern computer technology. It is Kelly's difficulty index that is used in this study.

OTHER PROCESS TESTS

A number of tests have become available in the past decade which have been addressed specifically to the task of assessing process ability. Most of the developers used rather traditional methods of test development: a pool of test items is generated, a panel of qualified judges examine the validity of the items and inappropriate items are dropped from the pool. The surviving validated items are then tried out on a sample of subjects similar to the target population. Item analysis data on the items is obtained and poor items are either revised or dropped from the pool. The resulting items make up the

test which is usually normed by administration to a fairly large sample of the target population.

Cooley and Klopfer⁵¹ in their development of the Test on Understanding Science (TOUS) added an additional evidence of validity by administering TOUS as a pre- and post-test to a group of talented students who spent a summer working with scientists. Their scores improved. Whether or not this can be interpreted as evidence that TOUS measures processes may be questioned. One might argue that TOUS is only measuring factual recall and that their factual knowledge increased as a result of their experiences.

Welch and Pella⁵² sought additional evidence of validity for their test, The Science Process Inventory (SPI), by administering it to students, teachers, and scientists. They suggested that since scientists would be expected to know the most and students the least about science processes, the fact that the scientists obtained the highest and the students obtained the lowest mean score on SPI, this was evidence of validity. A comparison of the above ranking with their ranking on a test that claimed to measure only factual recall might have been interesting.

Tannenbaum⁵³ developed the Test of Science Processes (TOSP). He recognized the difficulty of validating a process test. In addition to validation by expert opinion, he asked the teacher of one group to rank his students in order according to their process ability. He then used the correlation between the teachers' ranking and the students' scores on his test as evidence of validity. In order to reduce the reading required, TOSP contains some black and white pictures printed in the test booklet and some color slides which must be projected.

Beard⁵⁴ also has constructed a process test with the claim to validity based on the opinion of a panel of judges. In this case an attempt was made to minimize the reading required by synchronizing a taped script with color slides.

Morgan⁵⁵ developed the Science Test for Evaluation of Process Skills (STEPS). Again, expert opinion was the source of the claim of validity. In this case the reading problem was minimized through the use of film loops. One loop was used for each of the five sections of the test.

Ebel⁵⁶ has supported the use of pictures as a means of reducing the reading requirement, especially in contexts where a great deal of explanation would otherwise be required to set the task. However, when the pictures are projected for the entire group this constitutes pacing, which would seem to have many of the adverse effects of a speeded test as discussed earlier in this review. The preceding considerations prompted the use of printed pictures for TSPT.

Probably the most convincing claim of test validity is made by the Individual Competency Measures developed to accompany the SAPA program.⁵⁷ It uses the same materials and contexts in testing the ability to use the processes as it uses to teach them. Thus if the processes are taught by the SAPA program, ability to use them can reasonably be expected to be tested by the Individual Competency Measures. In spite of their strong claim of validity, the Individual Competency Measures are not extensively used. The reason is their low time efficiency. The Individual Competency Measures, by their very nature, require an individualized testing situation. They also frequently require that equipment and materials be available for the

student to manipulate as part of the evaluation.⁵⁸ This tends to make them even less attractive to the testor.

Nelson⁵⁹ has developed a test, the Inquiry Skills Measures (ISM) very similar to the Individual Competency Measures in that it is administered on an individualized basis and requires that materials be available for use as part of the test. Again, time efficiency detracts from the utility of the test.

THE NEED FOR EXTERNAL CRITERION REFERENCED VALIDATION

Over the years a number of authorities have expressed concern over the methods of test construction traditionally used. Buross⁶⁰ warned that to develop a valid test instrument, items should not be selected based on their correlation with the total test score since the entire test may prove to be invalid. Findley⁶¹ has warned of the danger inherent in the use of "expert opinion" as a means of validation. Barclay⁶² has written, "... the difficulty with testing usage centers very much on the determination of an adequate criterion which is independent of the testing instrument."

Since the purpose of process testing is to get at children's ability to think and reason, and since it is difficult to know how a child arrives at a given response, it seems reasonable that the matter of validity deserves special attention in the case of the process test. Horrocks⁶³ has indicated that due to the difficulty of writing good process test items perhaps the greatest hazard in testing for the processes is that of not developing valid items. The concern for item validity is a primary concern of this study.

THE WORK OF FYFFE AND ROBISON

Fyffe⁶⁴ and Robison⁶⁵ approached the problem of validation by recognizing that perhaps the Individual Competency Measures with their previously mentioned claim of validity represented Ebel's "clearly superior" test. But in this case there is a need for another test because of the time efficiency problem which limits the usefulness of the Individual Competency Measures. Therefore their procedure was as follows:^{66,67}

They began by first selecting a representative sample of the Individual Competency Measures to be used as their external criterion. The next step was to prepare multiple choice test items. Ideas for items were drawn from a review of textbooks and laboratory manuals. In order to assure face validity, a committee of judges composed of both faculty and graduate students reviewed the items using the following procedure:^{68,69}

"The procedure followed in the review of items was to provide each reviewer with a list of the objectives for the four integrated processes at the same time that proposed test items were available. Each test item was then identified as measuring one objective or skill for a particular process. The reviewer then had two considerations to decide: (1) Does the item require the use of the specific process skill identified? and (2) Does the item present enough information that a skillful seventh grade student can respond correctly?"

External Criterion Referenced Validation

Fyffe and Robison administered the selected Individual Competency Measures and then their items to a group of subjects. They used the scores on the Individual Competency Measures to obtain the "upper 27 percent" and the "lower 27 percent" groups needed for the discrimination calculation in the item analysis. Thus decisions could be made about

the value of the items based not on the test under construction and not based on "expert opinion," but rather, based on the ability of the item to discriminate between students who did well (upper 27 percent) and those who did poorly (lower 27 percent) on the external criterion, the Individual Competency Measures. This procedure should satisfy the concern expressed in the preceding section, provided the Individual Competency Measures can be accepted as defining what is meant by the science processes.

Weakness

Fyffe reports, "Many of the items for the two processes of interest were pre-tested on two seventh grade students."⁷⁰ This is the extent of their item tryout before entering into the major portion of their study which involved administration of the Individual Competency Measures and their items to 56 students. An examination of the item analysis data they obtained from this administration of their items^{71,72} reveals that a number of their items (i.e., items 9, 10, 17, 18, 21) could probably have been improved by item tryout and revision.

SUMMARY

In this chapter, the trend toward process teaching and the difficulties this shift in emphasis has posed for testing was presented. Some of the issues surrounding the mechanics of testing were briefly examined, including speeded vs power tests, the use of a blueprint, the optimum number of alternatives, whether item order should be a concern, what value of item difficulty should be used, and finally, a brief examination of the discrimination index. A short survey of some of the recent tests which have been developed for the purpose of

assessing science processes was presented and the method of validation for each test was examined.

The concern which authorities in testing have expressed over the traditional methods used in test construction was examined, especially as these relate to the validation of tests which attempt to assess process abilities. Finally, the work of Fyffe and Robison with their use of external criterion referenced validation was reviewed.

FOOTNOTES

¹National Education Association, The Central Purpose of American Education (Washington, D.C., 1961), p. 19.

²Henry P. Cole, "Process Curricula and Creativity Development," Journal of Creative Behavior, 3, 253 (Fall 1969).

³Terry Borton, "What's Left When School's Forgotten?" Saturday Review, 53, 69-71, 79 (April 18, 1970).

⁴American Association for the Advancement of Science, The Psychological Bases of Science - A Process Approach, AAAS Misc. Publication (1965), pp. 65-68.

⁵American Association for the Advancement of Science, Science - A Process Approach Commentary for Teachers, AAAS Misc. Publication, 68-7 (1968).

⁶Ibid, p. 163.

⁷Joint Committee of the American Association of School Administrators, Testing, Testing, Testing (Washington, D.C.: American Association of School Administrators, 1962), p. 9.

⁸Robert L. Ebel, "The Relation of Testing Programs to Educational Goals," The Sixty-second Yearbook of the National Society for the Study of Education (University of Chicago Press, Chicago, 1963).

⁹Progress Report of the Panel on Educational Research and Development, Innovation and Experimentation in Education (U.S. Government Printing Office, Washington, D.C., 1964), p. 44.

¹⁰Ralph W. Tyler, "Resources, Models, and Theory in the Improvement of Research in Science Education," Journal of Research in Science Teaching, 5, 43 (1967).

¹¹Eugene Lee, New Developments in Science Teaching (Wadsworth Press, Belmont, California, 1967), p. 69.

¹²Louis Kuslan and A. H. Stone, Teaching Children Science: An Inquiry Approach (Wadsworth Press, Belmont, California, 1968), p. 228.

¹³Richard B. Smith, "Approach to Measurement in the New Science Curriculum," Science Education, 53, 411-415 (December 1969).

¹⁴John R. Bormuth, On the Theory of Achievement Test Items, (Chicago: University of Chicago Press, 1970).

¹⁵Richard C. Anderson, "How to Construct Achievement Tests to Assess Comprehension," Review of Educational Research, 42, 145-170 (1972).

¹⁶L. Lisonbee, "Testing, What For?" Science Teacher, 33, 27-29 (May 1966).

¹⁷Hulda Grobman, Evaluation Activities of Curriculum Projects, AERA Monograph Series on Curriculum Evaluation, No. 2 (Chicago: Rand McNally, 1968).

¹⁸Robert H. Ennis, "Needed: Research in Critical Thinking," Educational Leadership, 21, 17-20, 39 (October 1963).

¹⁹Robert E. Stake and T. Denny, "Needed Concepts and Techniques for Utilizing More Fully the Potential of Evaluation," The Sixty-eighth Yearbook of the National Society for the Study of Education, 2 (Chicago: University of Chicago Press, 1969).

²⁰Ralph W. Tyler, "Educational Evaluation: New Roles, New Means," The Sixty-eighth Yearbook of the National Society for the Study of Education, 2 (Chicago: University of Chicago Press, 1969).

²¹Hulda Grobman, "Curriculum Development and Evaluation," Journal of Educational Research, 64, 436-442 (July 1971).

²²Banesh Hoffman, The Tyranny of Testing (Crowell-Collier Press, New York, 1962).

²³Ralph W. Tyler, Basic Principles of Curriculum and Instruction (Chicago: University of Chicago Press, 1950).

²⁴R. Thorndike and Hagen, Measurement and Evaluation in Psychology and Education (New York: John Wiley, 1955).

²⁵F. B. Davis, Educational Measurements and Their Interpretation (Wadsworth, Belmont, California, 1964).

²⁶Robert L. Ebel, Essentials of Educational Measurement (Prentice-Hall, Inc., Engelwood Cliffs, New Jersey, 1972).

²⁷C. Terranova, "Relationship Between Test Scores and Test Time," Journal of Experimental Education, 40, 81-83 (Spring 1972).

²⁸Ross E. Traut and R. K. Hambleton, "The Effect of Scoring Instructions and Degree of Speededness on the Validity and Reliability of Multiple-Choice Tests," Educational and Psychological Measurement, 32, 737-758 (1972).

²⁹Ebel, op. cit., p. 108.

³⁰Franklin R. Evans and R. R. Reilly, "A Study of Speededness as a Source of Test Bias," Journal of Educational Measurement, 9, 123-131 (Summer 1972).

³¹Robert M. W. Travers, How to Make Achievement Tests (New York: Odyssey Press, 1950), p. 25.

³²E. F. Lindquist, ed., Educational Measurement (American Council on Education, Washington, D.C., 1951), pp. 119-495.

³³Ebel, op. cit., p. 364.

³⁴A. Tversky, "On the Optimal Number of Alternatives at a Choice Point," Journal of Mathematical Psychology, 1, 386-391 (1964).

³⁵Frank Costin, "Optimal Number of Alternatives in Multiple-Choice Achievement Tests: Some Empirical Evidence for a Mathematical Proof," Educational and Psychological Measurements, 30, 353-358 (Summer 1970).

³⁶Robert L. Ebel, "Expected Reliability as a Function of Choices Per Item," Educational and Psychological Measurement, 29, 565-570 (1969).

³⁷Ronald L. Flaugher, R. S. Melton, and C. T. Myers, "Item Rearrangement Under Typical Test Conditions," Educational and Psychological Measurement, 28, 813-824 (Autumn 1968).

³⁸C. Munz and A. D. Smouse, "Interaction Effects of Item-Difficulty Sequency and Achievement-Anxiety Reaction on Academic Performance," Journal of Educational Psychology, 59, 370-374 (October 1968).

³⁹Marshall H. Brenner, "Test Difficulty, Reliability, and Discrimination as Functions of Item Difficulty Order," Journal of Applied Psychology, 48, 98-100 (April 1964).

⁴⁰Ronald N. Marso, "Test Item Arrangement, Testing Time, and Performance," Journal of Educational Measurement, 7, 113-118 (Summer 1970).

⁴¹Naomi C. Klosner and E. K. Gellman, "The Effect of Item Arrangement on Classroom Test Performance: Implications for Content Validity," Educational and Psychological Measurement, 33, 413-418 (1973).

⁴²P. M. Symonds, "Factors Influencing Test Reliability," Journal of Educational Psychology, 19, 73-87 (1938).

⁴³Fredrick B. Davis, "Item Analysis in Relation to Educational and Psychological Testing," Psychological Bulletin, 49, 97-121 (1952).

⁴⁴J. F. Adams, "Test Item Difficulty and the Reliability of Item Analysis Methods," Journal of Psychology, 49, 255-262 (1960).

⁴⁵J. C. Wofford and T. L. Willoughby, "The Effects of Test Construction Variables Upon Test Reliability and Validity," California Journal of Educational Research, 20, 96-106 (May 1969).

⁴⁶Frederick B. Davis, 1971 AERA Conference Summaries: II. Criterion Referenced Measurement (ERIC Clearinghouse on Tests, Measurement and Evaluation, Princeton, New Jersey, 1972).

⁴⁷Truman L. Kelly, "The Selection of Upper and Lower Groups for the Validation of Test Items," Journal of Educational Psychology, 30, 17-24 (1939).

⁴⁸L. S. Feld, "Note on Use of Extreme Criterion Groups in Item Discrimination Analysis," Psychometrika, 28, 97-104 (1963).

⁴⁹Wofford and Willoughby, loc. cit.

⁵⁰Max D. Engelhart, "A Comparison of Several Item Discrimination Indices," Journal of Educational Measurement, 2, 69-76 (June 1965).

⁵¹William W. Cooley and L. E. Klopfer, "The Evaluation of Specific Educational Innovations," Journal of Research in Science Teaching, 1, 73-80 (1963).

⁵²Wayne W. Welch and M. O. Pella, "The Development of an Instrument for Inventorying Knowledge of the Processes of Science," Journal of Research in Science Teaching, 5, 64-68 (1967).

⁵³R. S. Tannenbaum, "Development of the Test of Science Processes," Journal of Research in Science Teaching, 8, 123-136 (1971).

⁵⁴Jean Beard, "The Development of Group Achievement Tests for Two Basic Processes of AAAS Science - A Process Approach," Journal of Research in Science Teaching, 8, 179-183 (1971).

⁵⁵D. A. Morgan, "STEPS Science Test for Evaluation of Process Skills," The Science Teacher, 38, 77-79 (November 1971).

⁵⁶Robert L. Ebel, "Writing the Test Item," Educational Measurement, E. F. Lindquist, ed. (American Council on Education, Washington, D.C., 1951), pp. 185-249.

⁵⁷H. Walbesser, "Science Curriculum Evaluation: Observations on a Position," The Science Teacher, 33, 34-39 (February 1966).

⁵⁸American Association for the Advancement of Science, An Evaluation Model and Its Application, Second Report (AAAS, Washington, D.C., 1968), pp. 9, 10.

⁵⁹Miles A. Nelson and E. C. Abraham, "Inquiry Skill Measures," Journal of Research in Science Teaching, 10, 291-297 (1973).

⁶⁰Oscar K. Buros, "Criticisms of Commonly Used Methods of Validating Achievement Test Items," Proceedings of the 1948 Invitational Conference on Testing Problems, (Educational Testing Service, 1949), p. 18.

⁶¹Warren G. Findley, "Purposes of School Testing Programs and Their Efficient Development," The Sixty-second Yearbook of the National Society for the Study of Education (Chicago: University of Chicago Press, 1963), p. 8.

⁶²James R. Barclay, Controversial Issues in Testing (Boston: Houghton Mifflin, 1968), p. 60.

⁶³John E. Horrocks and T. I. Schoonover, Measurement for Teachers, (Charles E. Merrill Publishing Company, Columbus, Ohio, 1968), p. 70.

⁶⁴Darrel W. Fyffe, The Development of Test Items for the Integrated Science Processes: Formulating Hypotheses and Defining Operationally, Unpublished Doctoral Dissertation, Michigan State University (1971).

⁶⁵Richard Wayne Robison, The Development of Items which Assess the Processes of Controlling Variables and Interpreting Data, Unpublished Doctoral Dissertation, Michigan State University (1973).

⁶⁶Fyffe, op. cit., pp. 30-42.

⁶⁷Robison, op. cit., pp. 39-55.

⁶⁸Ibid, p. 41.

⁶⁹Fyffe, op. cit., p. 35.

⁷⁰Ibid, p. 36.

⁷¹Ibid, p. 107.

⁷²Robison, op. cit., p. 103.

CHAPTER III

PROCEDURE

This study consisted of several quite distinct phases which will be described independently. They are:

The Item Improvement Phase in which the items in the item pool were tried out and on the basis of the resulting item analysis data they were retained, edited or dropped from the pool.

The Validation Phase in which the items in the item pool were validated using the external criterion referenced validation method, and based on this validation the final form of TSPT form D was constructed. This phase of the study has many of the characteristics of a typical experimental study with hypotheses that are tested and either accepted or rejected based on statistical treatment of the data.

The Norming Phase in which TSPT form D was administered to a random sample of students and from the resulting data, norms were prepared for use with the test.

The final phase of the study consisted of the writing of the test manual in which TSPT form D is described, norming data is presented, and instructions for administration of the test and interpretation of the results are presented.

ITEM IMPROVEMENT

Construction of Form A

The items in the initial item pool developed by Fyffe and Robison were written with considerable attention to content so that there would

be reasonable assurance that they would assess the ability to use the science processes. Their procedure is described in Chapter II of this study. Their item analysis data provides considerable evidence that the lack of adequate item tryout and revision seriously limited the usefulness of their items as written. Thus the initial step in this study was to examine their items in the light of the item analysis data presented in their study.¹ Many of their items had to be rewritten and some of them were dropped from the pool as the result of the above item analysis. The criteria established for this and subsequent revisions during the item improvement phase of the study are:

1. The reading level of the item was kept within the ability of sixth grade students.

2. All alternatives were required to have been chosen.

3. Sufficient description preceded the item to set the task.

In some cases since more than one item was based on a given context, the group of items had to be included or excluded in toto.

4. The difficulty of the item (proportion of students missing the item) was required to be between 0.2 and 0.7.

5. The discrimination of the item as determined by conventional methods² was required to be at least 0.3. An empirical justification for the use of the 0.3 value is given in Appendix III-A.

Additional items were written to bring the item pool up to 80 items. Special care was taken in writing the new items to be sure that they were compatible in format, style, and language to those items written by Fyffe and Robison. The pictures required to clarify certain contexts were obtained, and the items were assembled and printed to make up TSPT form A Parts I and II, which is included in Appendix IV-A. The test

was separated into two parts as nearly equal in length as the contexts would permit since it was expected that 80 items would be too long a test to be administered in a typical class period and also probably too long for the children to handle in a single sitting as well.

For the tryout of TSPT form A two schools were contacted, one in Flint and one in Lansing, Michigan. Both schools were located in urban middle class neighborhoods. Each school contained two sixth grade classes with chance assignment of students to each class. Each school was recommended by their respective school district administrators as having a progressive science program. The important difference between them was that the Flint school used the SAPA program and the Lansing school used a traditional textbook oriented program. This method of sample selection was used in order to increase the likelihood that assumption 4 in Chapter I of this study was correct. In late May of 1973, TSPT form A was administered to one class in each school. In the traditional school two days elapsed between administration of Parts I and II. In the SAPA school, Parts I and II were given in the morning and afternoon, respectively, of the same day. Part I required about 30 minutes to complete and Part II required about 50 minutes. Probably Part II was too long for this age group. Although no time limit was imposed and no marked deterioration in their performance was detected, many students got quite restless before they finished Part II. The students marked their responses on spirit duplicated answer sheets. Their responses were transferred to machine readable forms. The Michigan State University test scoring service scored them and generated the usual test statistics and item analysis data.

Construction of Form B

Using the above item analysis data and the same revision criteria as used previously, many of the items of Form A were either rewritten or dropped from the item pool. The resulting 61 items made up TSPT form B which was also in two parts. Part I contained 33 items and Part II contained 28 items. In October of 1973 the cooperation of another pair of similar schools was sought for item tryout, one using the SAPA program and the other using a traditional science program. The SAPA school was in Flint and the traditional school was in Berrien Springs, Michigan. The schools used were very similar in size and socio-economic status to those used for the tryout of form A. In the SAPA school, Part I was administered the last period of the morning and Part II was administered the last period of the afternoon. In the traditional school, Parts I and II were administered on consecutive days. The students' responses were scored by the Andrews University computing center in Berrien Springs, Michigan. The usual item analysis and test statistics were produced.

Construction of Form C

Again applying the previously mentioned criteria a number of the items were rewritten but no more items were dropped from the pool. The resulting 61 items composed TSPT form C which was printed again in two parts of 33 and 28 items each. TSPT form C is included in Appendix IV-D.

In view of the small amount of revision required to produce form C, it was felt that TSPT was of sufficient quality to begin the next phase of the study.

VALIDATION PHASE

Design of the Study

At this point the study took on more of the characteristics of a classic research study describable using the notation of Campbell and Stanley³ as:

$$X \quad O_1 \quad O_2 \quad O_3$$

X: All the previous experiences of the validation sample.

O₁: The first observation consisting of the administration of the Individual Competency Measures.

O₂: The second observation consisting of the administration of TSPT form C.

O₃: The third observation consisting of the administration of the SRA test.

The sequence of this part of the study was:

1. Administration of the Individual Competency Measures.
2. Administration of TSPT form C.
3. Administration of the SRA test.
4. Scoring the above tests and analysis of the results.
5. Construction of TSPT form D, a revision of TSPT form C using the external criterion referenced method of test development as previously defined in this study.
6. Comparison of students' performance on the form D subtest of TSPT form C with their performance on the Individual Competency Measures and the SRA test.

Selection of Individual Competency Measures to be Used

An evaluation of the Individual Competency Measures in terms of their appropriateness to this study was conducted and Competency Measures were selected for use based on the following considerations:

1. Enough Individual Competency Measures were to be used to include at least 10 tasks on each of the 4 Integrated Processes.
2. The Individual Competency Measures used were to be representative of the total pool of Individual Competency Measures available for each Integrated Process.
3. The Individual Competency Measures used were not to require factual recall of a given activity or terminology within the SAPA program.
4. The tasks involved were required to fit the testing situation used and time span of approximately 5 minutes per Individual Competency Measure per child.

A listing of the Individual Competency Measures pool considered for use in this study with an indication of which were actually used is included in Appendix I-C of this study.

Sample Selection and Description

In October, 1973, the science consultant for the Flint, Michigan Community School System was requested to suggest a school in which the validation study could be conducted. The criteria for selection were:

1. The SAPA program was highly implemented in the school.
2. The school was "typical middle class" in all other respects.

The school recommended was the Pierce Community School in Flint, Michigan. It was located in a stable middle class suburban neighborhood that took pride in its school and was interested and involved in what it was doing. There were approximately 300 students enrolled in the first six grades with grade six composed of two classes of 29 and 26 students, respectively. The SAPA program had been used throughout the school for several years though none of the teachers had had any extensive training in the program. Due to absences at one time or another during the study,

the final sample size was reduced to 52 students. Since the same teacher taught science to both classes, no distinction was made in the study between them.

In a meeting with the principal and the sixth grade teachers a description of the study was presented and their cooperation was sought and obtained. Throughout the study the school personnel were extremely cooperative even though considerable disruption of their routine was unavoidable.

Facilities for Administering the Individual Competency Measures

Two rooms were used for administration of the Individual Competency Measures. Both were very small, but since the tests were individualized, this was no disadvantage. The room used in the mornings contained a long table which proved to be ideal for setting up the equipment used in some of the tests. The room used in the afternoons contained a sink which helped greatly for other tests. The testing rooms were near enough to the sixth grade classrooms so that little time was wasted in moving students from one room to another.

Administration of the Individual Competency Measures

The author administered all of the Individual Competency Measures to every child. Testing began on November 1 and continued through December 18, 1973. Every school day during this period was used. The total time required for testing was approximately two and one half hours per child. The time was divided into approximately 15 minute sessions in which three Individual Competency Measures were administered. The materials for testing three of the Individual Competency Measures were set up and all of the students were cycled through them before setting up the next three Individual Competency Measures. Each set up usually

required slightly over two days to complete. The teachers were very cooperative and allowed students to be called from their classrooms almost whenever they were needed. The children were always called in alphabetical order and it was not long before they could anticipate when they were to be called so that very little time was wasted and disruption within the classroom was minimized.

There was some concern that this procedure could produce some contamination due to children sharing with their friends information relative to test questions they knew they would be facing later. No ready method of avoiding the hazard was available. If such contamination did occur, it was not readily observable. Each situation seemed to be as unique to the last students to see it as it had been to the first.

Scores were obtained from the rating sheets prepared for each student on each Individual Competency Measure. One point was awarded for each task correctly done as indicated on the rating sheet. In most cases several tasks were involved for one Individual Competency Measure. The specific number for each Individual Competency Measure is recorded in Appendix I-C. These scores were analyzed and stored in the computer at the Andrews University computing center.

Administration of TSPT form C.

On December 19, 1973, the author administered TSPT form C to the validation sample. Part I was administered to both classes in the morning and Part II was administered in the afternoon. These responses were also stored in the computer at the Andrews University computing center.

Administration of the SRA Test

In January, 1974, the SRA tests were administered to all the students at Pierce Community School as part of their testing program.

Rather than wait for the results to be returned by SRA, the sixth grade students' responses were recorded by hand and these data were also stored in the computer at the Andrews University computing center.

Test Scoring

Since the Individual Competency Measures student responses were not of the multiple choice format, the students' scores and standard deviations on the total test and on each subtest were the only data obtained. On both of the other tests, which were multiple choice and therefore were amenable to conventional item analysis techniques, the students' responses together with the answer keys were stored in the computer. This allowed scoring and item analysis to be performed on any desired subtest at any time without the need to reenter any data. This method of data storage proved to be particularly advantageous in the construction of TSPT form D.

Construction of TSPT form D

With but minor exceptions, item improvement was complete with form C, all items having met readability and technical quality requirements. The important criteria imposed for this final revision involved questions of validity. The method of external criterion referenced validation as developed by Fyffe and Robison and described in Chapter I was applied as follows:

First the students were placed in rank order according to their performance on the Individual Competency Measures. The upper 27 percent of this group and the lower 27 percent of this group formed the "upper 27 percent" and the "lower 27 percent" respectively for the item analysis calculation which was performed on the TSPT form C items.

Next correlation coefficients were calculated for all items with the Individual Competency Measures scores. Based on the item's discrimination as calculated above and their correlation with the Individual Competency Measures scores, TSPT form D was constructed based on the requirement that both of the above indices have minimum values of 0.2 and the requirement that the item's context allowed its use. The reason for the use of the 0.2 minimum value is mentioned in Chapter I and empirical evidence that, at least for this study, 0.2 is appropriate is presented in Chapter IV, Table 6. Thirty-six items met the above requirements and were assembled to compose TSPT form D.

A machine scorable answer sheet was designed to be used with TSPT form D. Special care was exercised to make the answer sheet as easy to use as possible in order to minimize the confusion it would generate among children who had not used machine scorable answer sheets before. The printing and binding of these materials completed the construction of TSPT form D.

Hypotheses to be Tested

After both the Individual Competency Measures and TSPT form C had been administered to the validating sample, it was possible to test the following hypothesis:

1. Students' scores on each of the TSPT form C items will be more highly correlated with their scores on one of the four Integrated Processes than with their scores on any other process, and this correlation will indicate the subtest to which the item belongs. This hypothesis can be stated in the null form as follows:

H_0 : There are no differences at the .01 level of confidence among the correlations of the scores on each TSPT form C item

with the Individual Competency Measures subtest scores.

The directional alternate hypothesis is:

H_a : The score on each item of TSPT form C has a higher correlation with its score on one of the Individual Competency Measures subtests than with any of the other Individual Competency Measures subtest scores.

The real concern here is with the validity of item assignment. If a given item really assesses students' ability with respect to one of the science processes to a greater extent than any of the others, then that fact should be revealed by the correlation of students' scores on the item with their scores on the integrated processes subtests of the Individual Competency Measures. This correlation would add quantitative support to the "expert opinion" criticized in Chapter II of this study.

After the composition of TSPT form D was known, the validating sample scores on that subtest of TSPT form C were obtained. It then was possible to test the following hypothesis:

2. The scores on the TSPT form D subtest of TSPT form C will be more highly correlated with the scores on the Individual Competency Measures than with the scores on the SRA Science test. This hypothesis can be stated in the null form as follows:

H_0 : There is no difference at the 0.01 level of confidence between the correlation of the scores on the TSPT form D subtest of TSPT form C with the scores on the Individual Competency Measures and the correlation of the scores on the TSPT form D subtest of form C with the scores on the SRA Science Test.

The directional alternate hypothesis is:

H_a : The scores on TSPT form D subtest of TSPT form C are

more highly correlated with the scores on the Individual Competency Measures than with the scores on the SRA Science test.

Again the concern is that of validity. It is claimed that the Individual Competency Measures assess process ability.⁴ It is claimed, as has been mentioned, that the SRA Science test assesses "mainly factual knowledge."⁵ If these judgments are correct, then a comparison of the correlations of TSPT scores with scores on these tests should provide quantitative evidence for the validity of TSPT.

Testing Hypothesis 1

The correlation between the validating sample scores on TSPT form C items and Individual Competency Measures subtests were calculated and significant differences among those correlations were sought. Intercorrelations among the Individual Competency Measures subtests were also computed and significant differences among them were sought using Hostellings t test for significance of differences of correlated correlations.

Testing Hypothesis 2

After TSPT form D was constructed the correlation between the validating sample scores on the form D subtest of TSPT form C was calculated. The correlation between the TSPT form D scores and the SRA Science test scores was also calculated and a t test for the significance of the difference was performed.

MULTIPLE REGRESSION ANALYSIS

Finally, to elucidate the relations among the variables, the validating sample scores on the Individual Competency Measures, TSPT,

SRA Science, and SRA Reading were taken to the Michigan State University Computing Center where a multiple regression was performed. The Individual Competency Measures score was the dependent variable and the TSPT form C, SRA Science and SRA Reading scores were the independent variables.

NORMING TSPT FORM D

The next phase of the development of TSPT consisted of generation of norming information in order that potential users of TSPT will have a frame of reference from which to judge how TSPT might perform in their situation.

Sample Selection

In order to restrict the travel required, the population from which the norming sample was drawn consisted of the public schools within a 50 mile radius of Berrien Springs, Michigan that contained sixth grade classes as listed in the 1973-74 Michigan and Indiana state school directories. There were 243 schools in this population.

The schools were numbered consecutively from 1 through 243. The first 20 of a set of computer generated random integers from 1 through 243 were obtained. The schools assigned these numbers made up the norming sample. Of these schools, one refused to cooperate, claiming that accountability studies, federal funding, and other activities were imposing too heavy a testing program to allow any more. The resulting norming sample consisted of 19 schools.

Data Collection and Reduction

Testing the norming sample was begun in early April of 1974 and was completed in late May. Contact with the schools was first made

through the school superintendent. A brief description of the development of TSPT, the way their school was selected to participate, and the extent of their involvement was given. If the school system was small, the superintendent frequently gave immediate permission to contact the principal. If it was large, referral was usually made to the science or testing consultant who frequently wished to confer with the principal before giving permission.

The next step was to contact the school principal to set up an appointment to meet personally with him and his sixth grade science staff. At this meeting the need for process tests, the development of TSPT, the method of selecting the norming sample, and the part they were being asked to play was outlined. In most cases the school personnel were very willing to cooperate. A date was then agreed upon for administration of TSPT and a form was completed indicating size and number of classes, name of teachers, and science program presently being used. A set of directions for administering TSPT was given to the teacher (included in Appendix III-B) and these were reviewed briefly with him.

On the morning of the date TSPT was to be given, the tests and answer sheets were delivered to the school office. The completed tests were picked up either the same afternoon or the next morning.

The students' answer sheets were checked to see that the name block was correctly filled in and the responses were properly marked. They were then delivered to the Andrews University computing center for reading. A printout of student responses was checked against their answer sheets to correct any reading errors that were made. The Opscan 100 reader used was remarkably forgiving of sixth grade students ability to stay within the proper response field, mark plainly, and erase cleanly. Of the over

46,800 responses read less than 20 errors were detected. After making any needed corrections, the tests were scored and the following information from each school was stored in the computer: The school identification, the students' responses and scores, the school mean and standard deviation.

Feedback was sent to the schools in two parts. A computer printout of students' names and scores together with the number of items on the test, number of subjects who took the test, mean score, standard deviation, mean difficulty, mean discrimination, KR20 reliability, and standard error of measurement were returned to the schools within a few days after they took the test. Following completion of testing, another letter was sent to the norming sample schools which contained a computer printout of the frequency distribution for the entire sample of 1301 students and the test statistics as listed above for the total sample together with a frequency distribution of the school means.

TEST MANUAL PREPARATION

The final step in the development of TSPT was the preparation of the test manual. A brief description of the development of TSPT is presented first. Then the method of norming the test and a presentation of the norming statistics and frequency distribution is included. Finally, instructions for using TSPT and interpreting the results complete the test manual. A copy of the test manual is presented in Appendix III-C.

SUMMARY

This study may conveniently be broken down into the following phases:

Item Improvement

The items developed by Fyffe and Robison were revised using the data from their study. Additional items were added to the pool and two more item tryout and revision cycles were carried out using conventional item analysis procedures. The result was TSPT form C.

Validation

The Individual Competency Measures of SAPA, TSPT form C, and the SRA test were all administered to the validation sample. The correlation of each TSPT form C item score with each Individual Competency Measures subtest score was calculated to test hypothesis one, that TSPT form C items could be objectively placed in the appropriate Integrated Process subscale.

TSPT form D was constructed using the external criterion referenced validation method of test development which uses student performance on the Individual Competency Measures as the criterion for selecting items from TSPT form C for inclusion in TSPT form D.

The correlation of TSPT form D scores with the Individual Competency Measures scores was calculated. Hypothesis two that TSPT form D scores were more highly correlated with the Individual Competency Measures scores than with the SRA Science test scores was tested.

Finally, in an effort to more fully elucidate the relationships among the various tests, a multiple regression analysis was performed using the Individual Competency Measures scores as the dependent variable and TSPT, SRA Science, and SRA Reading scores as independent variables.

Norming

TSPT form D was administered to a random sample of 1301 sixth grade students for the purpose of obtaining norming data for TSPT form D.

Test Manual Preparation

Finally a test manual was prepared for TSPT form D containing a brief sketch of the development of TSPT form D, the norming data, and instructions for use of TSPT form D.

FOOTNOTES

¹Darrell W. Fyffe, The Development of Test Items for the Integrated Science Processes: Formulating Hypotheses and Defining Operationally, Unpublished Doctoral Dissertation, Michigan State University (1971), pp. 98-116.

²Robert L. Ebel, Essentials of Educational Measurement (Prentice-Hall, Englewood Cliffs, New Jersey, 1972), pp. 388-401.

³Donald T. Campbell and J. C. Stanley, Experimental and Quasi-Experimental Designs for Research (Chicago: Rand McNally, 1963).

⁴G. Billings, "Cognitive Levels of Elementary Science Tests," School Science and Mathematics, 71, 824-830 (December 1971).

⁵Clarence H. Nelson. "Review - Science Research Associates Achievement Series: blue version," in Buros, Seventh Mental Measurement Yearbook, 2 (Gryphon Press, Highland Park, New Jersey, 1972), pp. 1231-33.

CHAPTER IV

ANALYSIS OF RESULTS

The purpose of this study was to begin with the test items developed by Fyffe and Robison and to develop a test, TSPT, designed to assess students' ability to use the science processes as identified by the SAPA program. The study can quite naturally be divided into several phases. The results from each phase will be analyzed in this chapter.

ITEM IMPROVEMENT

Construction of Form A

After applying the revision criteria presented in Chapter III to the items and data available in Fyffe and Robison's study, many of their items were rewritten, some were dropped, and additional items were added to the pool. The result was TSPT form A which is included in Appendix IV-A.

Logical analysis of the items based on the contexts identified by SAPA indicated the assignment of items to the Integrated Processes as presented in Table 1, under the column heading marked "form A." Appendix IV-B contains the subject assignments for each item.

TABLE 1
ITEM SUBTEST ASSIGNMENTS

<u>Process</u>	<u>Number of Items</u>		
	<u>form A</u>	<u>form B</u>	<u>form C</u>
Interpreting Data	24	22	22
Controlling Variables	15	10	10
Formulating Hypotheses	18	11	11
Defining Operationally	<u>23</u>	<u>18</u>	<u>18</u>
Total	80	61	61

Results of the Tryout of Form A

TSPT form A was tried out in a classroom where the SAPA program was used and in a classroom where a traditional science program was used. The item analysis data from the tryout of TSPT form A are recorded in Appendix IV-C. The rest of the test statistics for TSPT form A are presented under the "form A" heading of Table 2. The purpose of this tryout was to obtain item analysis data for use in revising the test items, but since data was obtained from both a SAPA school and a traditional school, comparisons of the two schools' performance are possible.

A t test of significance of the differences between the SAPA and Traditional (Trad.) mean scores reveals that the mean scores for SAPA and Traditional students are not significantly different at the 0.01 level.

TABLE 2
TSPT TEST STATISTICS

	Form A		Form B		Form C	Form D	
	<u>SAPA</u>	<u>Trad.</u>	<u>SAPA</u>	<u>Trad.</u>	<u>SAPA</u>	<u>Form C Subtest</u>	<u>Norming Sample</u>
Number of Subjects	32	32	31	21	52	52	1301
Number of Items	80	80	61	61	61	35	36
Mean Score	32.45	26.99	27.65	28.95	27.12	17.12	17.9
Std. Deviation	8.10	10.39	7.75	8.31	9.42	7.75	6.90
KR20 Reliability	0.70	0.76	0.77	0.82	0.76	0.89	0.84
Std. Error	4.44	5.09	3.73	3.49	3.55	2.56	2.69
Mean Difficulty	0.59	0.66	0.56	0.53	0.56	0.51	0.50
Mean Discrimination	0.31	0.33	0.34	0.32	0.38	0.56	0.50

SAPA: Science program used in the school was SAPA.

Trad.: Science program used in the school was Traditional Textbook oriented.

Both schools had standardized test scores available so comparisons among these tests and TSPT form A were possible. These are presented in Table 3.

TABLE 3
TSPT FORM A CORRELATION TABLE

		Pearsons Product Moment Correlation
Traditional School	Form A with SAT Science	0.45 *
	Form A with SAT Reading	0.51 **
	SAT Science with SAT Reading	0.86 **
SAPA School	Form A with SRA Science	0.74 *
	Form A with SRA Reading	0.70 **
	SRA Science with SRA Reading	0.91 **
* Significant at 0.01 level. ** Significant at 0.001 level. SAT - Stanford Aptitude Test SRA - Science Research Associates Test		

Since form A is the preliminary form of TSPT no very great importance should be attached to these results but they do lend support to later work.

Results of the Tryout of Form B

The revision criteria as listed in Chapter III in the section titled ITEM IMPROVEMENT were applied to the data obtained from the

tryout of TSPT form A. A number of items were dropped and others were rewritten. The form A items which survived and were included in form C are indicated in Appendix IV-E. Since TSPT forms B and C are quite similar, form B is not presented; however, by comparing forms A and C presented in Appendices IV-A and IV-D respectively, a good idea of this phase of the revision process can be obtained. TSPT form B was tried out in the same manner as form A. The test statistics for TSPT form B are presented under the heading "form B" in Table 2.

A t test for significance of the differences between the SAPA and traditional mean scores reveals again no significant difference.

Form B is also a preliminary form, so again, no great importance should be attached to the data from it, but they do show that revision has improved the test.

Results of the Tryout of Form C

Applying the revision criteria again produced mostly small revisions with no items being dropped from the test. The result was TSPT form C which is included in Appendix IV-D.

Form C was administered to the Validating Sample which contained only sixth grade children who were being taught science using the SAPA materials. This sample is more fully described in Chapter III. The item analysis data is included in Appendix IV-E and the Validation Sample scores on TSPT form C are presented in the second column of Appendix IV-F. A plot of these data reveals a very slight positive skew. The other test data are presented under the heading labeled "form C" in Table 2.

The Integrated Processes subscales of form C are all highly correlated with one another and with the total test. These correlations

are presented in Table 4. The decimal points are suppressed.

TABLE 4
TSPT FORM C SUBTEST CORRELATIONS

	I	II	III	IV
I Interpreting Data	(87)			
II Controlling Variables	59	(78)		
III Formulating Hypotheses	59	53	(77)	
IV Defining Operationally	63	57	52	(86)

The values in parentheses are the correlation of each subtest with the total test. All of these correlations are significant at the 0.001 level.

VALIDATION

The Individual Competency Measures

The Individual Competency Measures previously identified and enumerated in Appendix I-C were administered to the validation sample. The total scores are very slightly negatively skewed with the Interpreting Data and Controlling Variables subtests accounting for most of the skew. These results are also displayed in Appendix IV-F.

Testing Hypothesis 1

The null form of hypothesis 1 can be stated as follows:

H_0 : There are no differences at the 0.01 level of confidence among the correlations of the validating sample scores on each TSPT form C item with their scores on each of the Individual

Competency Measures subtests of Interpreting Data, Controlling Variables, Formulating Hypotheses, and Defining Operationally. Written symbolically:

$$R_{I,ID} = R_{I,CV} = R_{I,FH} = R_{I,DO} \text{ for all } I$$

$R_{I,ID}$: The correlation of the validation sample scores on the TSPT item (I) with their scores on the Interpreting Data subtest of the Individual Competency Measures.

$R_{I,CV}$: The above correlation with the Controlling Variables subtest.

$R_{I,FH}$: The above correlation with the Formulating Hypotheses subtest.

$R_{I,DO}$: The above correlation with the Defining Operationally subtest.

In order to test the above hypothesis, the correlation between the Validation Sample scores on each TSPT form C item and their scores on each of the Individual Competency Measures subtests was computed. The resulting correlations were then tested for significance of differences using a *t* test for correlated correlations. The results are presented in Appendix IV-G. Of the 366 *t* tests performed, the null hypothesis was rejected only six times. In no case were the differences sufficient to unambiguously assign the item to one and only one of the subtests at the 0.01 level. At the 0.1 level this procedure unambiguously assigned four items to one and only one of the Individual Competency Measures subtests. These four items, their assignments based on the above correlations and their logical assignments based on their agreement with the SAPA contexts are presented in Table 5.

TABLE 5
TSPT FORM C ITEM ASSIGNMENTS

<u>Item Number</u>	<u>Assignments</u>	
	<u>Correlation</u>	<u>Logic</u>
40	ID	FH
55	DO	ID
56	CV	DO
61	FH	ID

Based on the above data the hypothesis that the Integrated Process which a given TSPT form C item assesses could be objectively determined based on the correlation of the Validation Samples scores on the item with their scores on the Individual Competency Measures subtests was not supported and the items' assumed relation to the integrated processes was not used as a criterion for selection of form C items to be included in TSPT form D. The question of these subtests is considered further in the discussion section of this chapter.

TSPT form D

The minimum level of external criterion referenced discrimination that should be required for inclusion of an item in TSPT form D was empirically tested by constructing a number of subtests of TSPT form C and examining their statistics. The results are presented in Table 6.

TABLE 6
TSPT FORM D ITEM SELECTION CRITERIA

	Form C	<u>Minimum Discrimination</u>				Form D
		<u>.1</u>	<u>.2</u>	<u>.3</u>	<u>.4</u>	
Number of Items	61	47	40	27	16	35
Mean Score	27.12	22.21	19.38	13.17	8.25	17.12
Standard Deviation	9.42	9.14	8.54	6.61	4.35	7.75
KR20 Reliability	0.86	0.89	0.90	0.89	0.86	0.89
Standard Error	3.55	3.04	2.76	2.23	1.65	2.56
Mean Difficulty	0.55	0.53	0.52	0.51	0.48	0.51
Mean Discrimination	0.38	0.42	0.46	0.55	0.62	0.56
Correlation with the Individual Competency Measures	0.78	0.82	0.83	0.85	0.86	0.83

As expected, since all the statistics are using the same data, as the minimum discrimination level is raised, the correlation with the Individual Competency Measures and the mean discrimination go up and the mean difficulty and standard error go down, but the KR20 reliability seems to be greatest for a minimum discrimination of about 0.2. This may reflect the reduction in size of the test as the discrimination requirement is increased, but at any rate for this test, the minimum external criterion referenced discrimination of 0.2 as used by Fyffe¹ and Robison seems to be about right.

The item analysis data for TSPT form C using the external criterion referenced item analysis procedure described in Chapter I is presented in Appendix IV-E under the "ICM" heading.

Included in the last column of Table 6 are the test statistics for the form D subtest of TSPT form C using the Validation Sample data. It should be noted that one item which was actually used on TSPT form D (Item 8) is not included because a slight revision of this item produced a marked improvement in its performance. Form D, the final form of TSPT is presented in Appendix IV-H.

SRA Test

The results of the administration of the SRA Science and Reading tests to the Validation Sample are presented in the last two columns of Appendix IV-F. The test statistics are presented in Table 7. It should be noted that the items which are cross keyed by SRA as being on both the reading test and the science test are omitted.

TABLE 7
SRA TEST STATISTICS

	<u>Science</u>	<u>Reading</u>
Number of Items	40	60
Mean Score	25.82	40.7
Standard Deviation	8.45	12.86
KR20 Reliability	0.91	0.94
Standard Error	2.54	3.05
Mean Difficulty	0.35	0.32
Mean Discrimination	0.52	0.53

Testing Hypothesis 2

The null form of hypothesis 2 can be stated as:

H_0 : There is no difference between the correlation of the Validation Sample scores on TSPT form D subtest of TSPT form C with their scores on the Individual Competency Measures and the correlation of their scores on TSPT form D subtest of TSPT form C with their scores on the SRA Science test, or symbolically:

$$H_0: R_{TSPT, ICM} - R_{TSPT, SRAS} = 0$$

Where:

$R_{TSPT, ICM}$: The correlation of the Validating Sample scores on the TSPT form D subtest of TSPT form C with their scores on the Individual Competency Measures.

$R_{TSPT, SRAS}$: The correlation of the Validating Sample scores on the TSPT form D subtest of TSPT form C with their scores on the SRA Science test.

To test the above hypothesis a t test of significance of the difference between correlated correlations was performed. The correlations were obtained from the data presented in Appendix IV-F. The results are presented in Table 8. The t value obtained is not significant even at the 0.2 level. Thus the null hypothesis was not rejected.

TABLE 8

TSPT, ICM - TSPT, SRAS CORRELATION COMPARISON

N=52		
$R_{TSPT, ICM}$	=	0.83
$R_{TSPT, SRAS}$	=	0.79
$R_{ICM, SRAS}$	=	0.74
Significance of the difference $R_{TSPT, ICM} - R_{TSPT, SRAS}$:		
Calculated:	$t =$	0.76
For Significance (0.01, one tailed test):		
	$t =$	2.4

The Validating Sample scores on the form D subtest of TSPT form C correlate about as well with the SRA Science test scores as with the scores on the Individual Competency Measures at the 0.01 level. The lack of a significant difference is examined more fully in the discussion section below.

Discussion

Hypothesis 1: To elucidate the absence of significant differences among the TSPT item - Individual Competency Measures subscale correlations, the intercorrelation among the Integrated Processes subscales of the Individual Competency Measures were calculated. These are recorded in Table 9. A t test of significance of differences indicated no significant differences at the 0.01 level. Thus it can be argued that they are all measuring similar abilities and so it would be

very hard to find a test item that would correlate significantly higher with one subtest than with another.

TABLE 9
INDIVIDUAL COMPETENCY MEASURES
SUBTEST INTERCORRELATIONS

<u>Subtests</u>	<u>Correlations</u>
Interpreting Data - Controlling Variables	0.75
- Formulating Hypotheses	0.72
- Defining Operationally	0.65
Controlling Variables - Formulating Hypotheses	0.74
- Defining Operationally	0.65
Formulating Hypotheses - Defining Operationally	0.62

Perhaps one reason for the lack of significance differences among the subtests can be found in the definitions of the Integrated Processes as presented in Appendix I-B. For example, if by Interpreting Data SAPA means the ability to "...CONSTRUCT one or more inferences or hypotheses from a comparison of the information in two or more related tables...", should one be surprised to find the Interpreting Data and Formulating Hypotheses subtest scores highly correlated?

Hypothesis 2: The result of testing hypothesis 2 is unequivocal. It would have been the same even if the level of significance were changed by an order of magnitude. TSPT scores are as closely related to the SRA Science scores as they are to the Individual Competency Measures scores. To elucidate this result, the correlation of the Validating Sample scores on the Individual Competency Measures with

their scores on the SRA Science test was reexamined. For the sample size used, a correlation greater than about 0.5 is significant at the 0.001 level of confidence. The value of 0.74 previously reported for this correlation indicates that the Individual Competency Measures and the SRA Science tests are to a considerable extent measuring either the same or very closely related abilities. It is therefore very hard for a third test to be more highly correlated with one than with the other.

To further examine this result, a multiple regression was performed using the Individual Competency Measures scores as the dependent variable and TSPT form C, Individual Competency Measures, and the SRA Science test scores as independent variables. The results are presented in Table 10 and represented pictorially in Figure 1. An interesting feature of these data is that the SRA Reading test accounts for more of the Individual Competency Measures variance (65 percent) than any other test used. Another interesting feature is that when TSPT form C and the SRA Reading scores are taken together, they account for 72 percent of the Individual Competency Measures variance and completely overlap the SRA Science test so that adding the SRA Science test accounts for almost no additional variance. However, the important point in terms of hypothesis 2 is that the SRA Science test appears to be closely related to the Individual Competency Measures, accounting for nearly the same amount of variance as any other test used. This, as was previously mentioned, makes it very hard to construct a test that will be more closely correlated with the Individual Competency Measures than with the SRA Science test.

TABLE 10
MULTIPLE REGRESSION ANALYSIS

Dependent Variable: The Individual Competency
Measures Scores.

<u>Order of Inserting Independent Variables in the Equation</u>	<u>Total Variance Accounted for</u>
SRAS	0.60
SRAR	0.66
TSPT	0.72
SRAS	0.60
TSPT	0.70
SRAR	0.72
SRAR	0.65
SRAS	0.66
TSPT	0.72
SRAR	0.65
TSPT	0.72
SRAS	0.72
TSPT	0.64
SRAS	0.70
SRAR	0.72
TSPT	0.64
SRAR	0.72
SRAS	0.72

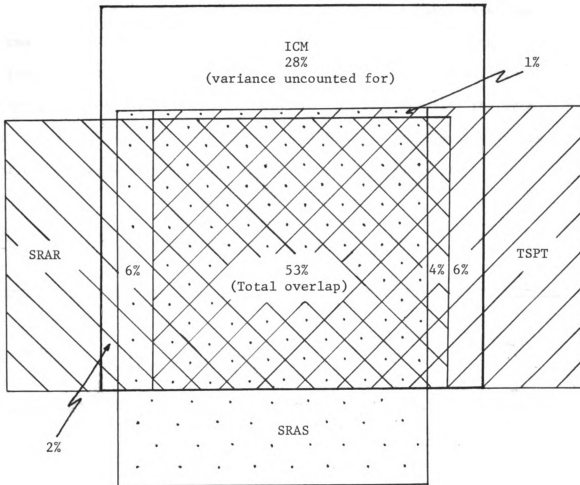


FIGURE 1

- ICM - The Individual Competency Measures, the dependent variable. Twenty-eight percent of the ICM variance is unaccounted for by any other test and 53 percent is accounted for by all the other tests.
- SRAR - The SRA Reading test, which accounts for 65 percent of the ICM variance. Two percent of the ICM variance is accounted for exclusively by this test.
- SRAS - The SRA Science test, which accounts for 60 percent of the ICM variance. None of the ICM variance is accounted for exclusively by this test.
- TSPT - TSPT form C, which accounts for 64 percent of the ICM variance. Six percent of the ICM variance is accounted for exclusively by this test.

NORMING TSPT FORM D

As was mentioned in Chapter III, the norming sample was drawn from public schools within a 50 mile radius of Berrien Springs, Michigan that contained sixth grade classes as listed in the 1973-74 Michigan and Indiana school directories. A map indicating the area included is presented in Appendix IV-I. Some of the characteristics of the population and the sample are revealed in Table 11.

TABLE 11
NORMING SCHOOLS CHARACTERISTICS

		<u>Population</u>	<u>Sample</u>
<u>Total Number of Schools Represented</u>		243	19
State	Michigan	119	9
	Indiana	124	10
Size of Community	25,000 or greater	118	10
	less than 25,000	125	9
Type of Community	Rural		3
	Inner City		4
Science Program	Innovative		3
	Textbook		5

The type of community and the type of science program categories in Table 11 should not be considered to be very precise classifications. They were derived strictly from intuitive impressions of the school, the classrooms, and the programs being conducted.

The statistics obtained from administration of TSPT form D to the norming sample are presented in Table 12. The complete item analysis is included in Appendix IV-J.

TABLE 12
TSPT FORM D TEST STATISTICS

Number of Items	36
Number of Subjects	1301
Median Score	17
Mean Score	17.900
Standard Deviation	6.899
KR20 Reliability	0.842
Standard Error	2.691
Mean Difficulty	0.503
Mean Discrimination	0.496

The frequency distribution of norming sample scores on TSPT form D is presented in Table 13. A plot of these data reveals a nearly normal distribution with a very slight positive skew.

TABLE 13
NORMING SAMPLE FREQUENCY DISTRIBUTION

<u>Score</u>	<u>Frequency</u>	<u>Standard Score</u>	<u>Percentile</u>
36	0	2.62	
35	0	2.48	
34	3	2.33	99.8
33	6	2.19	99.3
32	13	2.04	98.3
31	21	1.90	96.7
30	24	1.75	94.9
29	27	1.61	92.8
28	31	1.46	90.4
27	47	1.32	86.8
26	47	1.17	83.2
25	50	1.03	79.3
24	45	.88	75.9
23	58	.74	71.4
22	61	.59	66.7
21	53	.45	62.6
20	48	.30	59.0
19	51	.16	55.0
18	63	+ .01	50.2
17	59	- .13	45.7
16	55	.28	41.4
15	55	.42	37.2
14	72	.57	31.7
13	69	.71	26.4
12	65	.86	21.4
11	70	1.00	16.0
10	59	1.15	11.5
9	52	1.29	7.5
8	34	1.43	4.8
7	32	1.58	2.4
6	16	1.72	1.2
5	7	1.87	0.6
4	6	2.01	0.2
3	1	2.16	0.1
2	1	2.30	0.0
1	0	2.45	
0	0	-2.59	

SUMMARY

The significant correlation of TSPT scores with the Individual Competency Measures scores indicates that the external criterion referenced validation method of test construction is a fruitful approach to test construction.

The hypothesis that items could be objectively assigned to the ID, CV, FH, and DO subscales was not supported. These subscales within the Individual Competency Measures are so highly intercorrelated it was not possible to objectively assign a single test item unambiguously to any one subscale at the 0.01 level. Since the subscales were not objectively identifiable, no effort was made to identify them on TSPT.

One possible reason for not being able to identify these subscales is the overlap in their definitions.

The hypothesis that the correlation of TSPT scores with the Individual Competency Measures scores would be significantly higher than the correlation of TSPT scores with SRA Science test scores was not supported. One possible explanation for this outcome is the high correlation between the Individual Competency Measures scores and the SRA Science scores. This indicates that the two tests are very closely related in terms of what they measure. Thus it is very difficult for a third test to be more closely related to one than to the other.

The 0.01 level of significance was used in testing the hypotheses in this study.

FOOTNOTES

¹Darrell W. Fyffe, The Development of Test Items for the Integrated Science Processes: Formulating Hypotheses and Defining Operationally, Unpublished Doctoral Dissertation, Michigan State University (1971), p. 43.

CHAPTER V

SUMMARY AND CONCLUSIONS

Test development continues to be one of the major areas of concern in the educational community. Recent efforts to teach process abilities have intensified this concern. This study was an attempt to apply a method of test development which is different from the method widely used and to examine some of the properties of the resulting test.

SUMMARY

The method of test development used has been called by the descriptive name of External Criterion Referenced Validation. Simply stated, the procedure departs from standard test development practice in that the criterion for item acceptance is that the item discriminate between students who do well on the external criterion and students who do poorly on it.

For this study the external criterion is the Individual Competency Measures selected from the elementary school science program, Science - A Process Approach (SAPA), which defines what the test developed during this study, The Science Processes Test (TSPT), is intended to measure. The final form of TSPT consists of a 36 item, four alternative multiple choice test.

The item pool developed by Fyffe and Robison was used as the starting point for this study with additions, deletions, and revisions

being made based on three item tryouts using conventional item analysis procedures. The result was TSPT form C. Three tests were then administered to a single group of students, the Validation Sample. The tests were TSPT form C, the Individual Competency Measures which have been purported to measure ability to use science processes, and the SRA test which has been criticized as measuring "mainly factual knowledge." It was hypothesized that the items of TSPT form C could be assigned to the four subtests of the Individual Competency Measures based on the correlation of the students' scores on the item with their scores on the subtests. The data did not support this hypothesis. The subtests proved to be so highly intercorrelated that an item which correlated highly with one subtest was likely to correlate highly with one or more of the other subtests. Thus unambiguous item-subtest assignment based on item-subtest correlation did not occur for any of the TSPT form C items at the 0.01 level. Examination of the four instances when it did occur at the 0.1 level makes it hard to believe that they were more than chance occurrences. In view of their high intercorrelations no further reference was made to the Individual Competency Measures subtests, and no further attempts were made to distinguish among the supposedly different integrated processes as identified by SAPA.

TSPT form D was constructed using the external criterion referenced validation method of test development. The Validation Sample's performance on the Individual Competency Measures was used as the external criterion. The resulting test was highly correlated with the external criterion.

It was hypothesized that TSPT form D measures more of science processes than of factual knowledge and that students' scores on TSPT

form D should therefore correlate more highly with their Individual Competency Measures scores than with their SRA Science test scores. The data did not support this hypothesis. The SRA Science test was so highly correlated with the Individual Competency Measures that it seems unlikely that any test would show a significantly higher correlation with one than with the other and no significant differences were revealed by this study.

Finally norming data was obtained for use with TSPT form D by administering it to a random sample of 1301 sixth grade students and a test manual was prepared for the test.

CONCLUSIONS

The results of this study are quite clear cut. There is no temptation to talk about a test being "almost significant." The results would have been the same even if the levels of significance had been shifted either way by an order of magnitude. It seems unreasonable that the practical results of the study would have come out appreciably different if a different approach or data treatment had been used. In this respect the study was very satisfying and the following conclusions seem appropriate.

The Value of TSPT

Although the value of TSPT will only become apparent as it is used, its validity as indicated by the high correlation between scores on it and scores on the Individual Competency Measures, its quality as indicated by the test statistics, and its ease of administration and scoring all indicate that it should be of value to those concerned with testing.

The Value of External Criterion Referenced Validation

The external criterion referenced validation method of test development used to develop TSPT does provide a quantitative method for selecting test items which can be used as an objective alternate or supplement to the subjective judgment methods commonly employed. It seems reasonable that this method could be used to construct more time efficient tests in many situations where the most valid methods of evaluation involve more cumbersome methods such as direct observation, interviews, or other criterion performances.

Further, a quantitative estimate of the degree of validity may be inferred based on the correlation between scores on the test under development and scores on the criterion. Such correlations would give the potential user a more quantitative basis for making judgments relative to the merits of a given test than the rather qualitative reviewer's opinion on which the test user must currently rely.

Existence of Science Process Subscales

The four Integrated Processes as identified by SAPA are highly intercorrelated and as a result their existence as separately identifiable abilities is subject to question. The statements which define these processes and the Individual Competency Measures which assess students' ability to use them need to be made more distinct if they are to be objectively distinguishable.

Correlation Between Individual Competency Measures and SRA Scores

The high correlation between scores on the Individual Competency Measures and the SRA Science test can be interpreted as evidence that they measure much the same things. Although logical analysis by adult

experts may indicate that one test assesses "process ability" while another assesses "mainly factual knowledge," this appears to be no guarantee that in fact students' performances will vary widely from one test to the other.

IMPLICATIONS FOR FURTHER RESEARCH

Both the high intercorrelation of the Individual Competency Measures and the high correlation of process ability with factual knowledge observed in this study raise many questions. Do the Individual Competency Measures really measure process ability? If not, what are the processes and how can they be measured? Does the SRA Science test really measure mainly factual knowledge? If not, can such a test be constructed and can it then be objectively demonstrated that process ability and factual knowledge are indeed distinguishable one from the other?

The above mentioned high correlations could be explained if the subjects in fact largely lacked process ability. However, their scores on the Individual Competency Measures argue against such an interpretation. A more reasonable interpretation, and one which is not without support in the literature¹⁻⁴ is that process ability is rather ill defined and poorly understood. There is little real assurance for example that process ability and factual knowledge, which seem to be so different to test constructors and educational theorists, are indeed objectively and quantitatively distinguishable entities. Perhaps the most pressing need for further research is in this area. The processes need to be more precisely defined and distinctions both among the science processes and between process ability and factual knowledge need to be demonstrated. Further, assuming the above distinctions can

be demonstrated, the problem of the underlying psychological structures involved need to be elucidated. Until these problems are seriously addressed, the significance of the term "science processes" must be questioned. It is hoped that this test and the method of test development used here may be useful tools for attacking these problems.

Another way to interpret the observed high intercorrelations among the processes is to hypothesize that the children are taught the processes in a rather uniform fashion. This hypothesis could be tested in a number of ways. One way might be an experimental study in which children are intentionally taught the processes on a differential basis. A post test should reveal a lower intercorrelation among the processes for the experimental group and it should also reveal a significantly greater score improvement on the processes which were taught. The same sort of experiment could be performed with respect to the high correlation observed between process ability and factual knowledge. This kind of study would go a long way toward clarifying our understanding of what actually is involved in what goes by the rubric "process ability."

Correlational studies need to be performed on other pairs of tests, one of which has been classified as a process test and the other a factual knowledge test to see if they do indeed measure significantly different things. Studies of this nature should supplement, and perhaps eventually replace the subjective pronouncements of the reviewers on whose judgment we must now rely when we search for a test which measures "mainly processes" or "mainly factual knowledge." Studies of this kind will be helpful in moving testing out of the backwaters of educated guesswork and into the mainstream of scientific objective demonstration.

Finally, other investigators, perhaps in other fields, should experiment with the method of test construction used in this study. It would seem that to supplement and perhaps to ultimately replace, by a reliable objective method of item selection, the subjective judgment of "experts" used to establish the validity of test items is a goal worthy of the best efforts of test constructors. The technique used in this study could be a first step in this direction and should be tested and improved upon by other investigators. For example, a study could be performed in which the external criterion used to determine the validity of the test items could be, instead of the Individual Competency Measures as used in this study, direct observation of the children in the laboratory, the classroom, or even at play. Various techniques of this type could and should be tried which will promote a more objective scientific approach in the field of testing.

FOOTNOTES

¹Max D. Engelhart and John M. Beck, "The Improvement of Tests," The Sixty-second Yearbook of the National Society for the Study of Education (Chicago: University of Chicago Press, 1963).

²Julius M. Sassenrath, "The Factorial Composition of the Iowa Tests of Educational Development," California Journal of Educational Research, 16, 80-84 (March 1965).

³Stephen Klein, "Evaluating Tests in Terms of the Information They Provide," ERIC Document ED045, p. 699 (June 1970).

⁴Robert L. Ebel, Essentials of Educational Measurement (Prentice-Hall, Engelwood Cliffs, New Jersey, 1972), p. 109.

BIBLIOGRAPHY

BIBLIOGRAPHY

Books

- American Association for the Advancement of Science, An Evaluation Model and Its Application. Second Report. AAAS, Washington, D.C. (1968), pp. 9, 10.
- _____. Science - A Process Approach Commentary for Teachers. AAAS Misc. Publication 68-7, 1968.
- _____. The Psychological Bases of Science - A Process Approach. AAAS Misc. Publication 65-68, 1965.
- Barclay, James R., Controversial Issues in Testing. (Houghton Mifflin Co., Boston, 1968), p. 60.
- Bormuth, John R., On the Theory of Achievement Test Items. (Chicago: University of Chicago Press, 1970).
- Buros, Oscar K. "Criticisms of Commonly Used Methods of Validating Achievement Test Items," Proceedings of the 1948 Invitational Conference on Testing Problems (Educational Testing Service, 1949), p. 18.
- Davis, Frederick B. Educational Measurements and Their Interpretation. (Belmont, California: Wadsworth, 1964).
- _____. 1971 AERA Conference Summaries: II Criterion Referenced Measurement. (Princeton, New Jersey: ERIC Clearinghouse on Tests, Measurement and Evaluation, 1972).
- Ebel, Robert L. Essentials of Educational Measurement. (Engelwood Cliffs, New Jersey: Prentice-Hall, Inc., 1972).
- _____. "The Relation of Testing Programs to Educational Goals." The Sixty-Second Yearbook of the National Society for the Study of Education. (Chicago: University of Chicago Press, 1963).
- _____. "Writing the Test Item." E.F. Lindquist (ed). Educational Measurement, (American Council on Education, Washington, D.C., 1951), pp. 185-249.
- Engelhart, Max D., and John M. Beck. "The Improvement of Tests," The Sixty-second Yearbook of the National Society for the Study of Education. (Chicago: University of Chicago Press, 1963).

- Findley, Warren G. "Purposes of School Testing Programs and Their Efficient Development." The Sixty-second Yearbook of the National Society for the Study of Education. (Chicago: University of Chicago Press, 1963), p. 8.
- Grobman, Hulda. Evaluation Activities of Curriculum Projects, AERA Monograph Series on Curriculum Evaluation, No. 2. (Chicago: Rand McNally, 1968).
- Hoffman, Banesh. The Tyranny of Testing, (New York: Crowell-Collier Press, 1962).
- Horrocks, John E. and T. I. Schoonover. Measurement for Teachers, (Columbus, Ohio: Charles E. Merrill Publishing Co., 1968), p. 70.
- Innovation and Experimentation in Education, Progress Report of the Panel on Educational Research and Development. (U.S. Government Printing Office, Washington, D.C., 1964), p. 44.
- Joint Committee of the American Association of School Administrators. Testing, Testing, Testing. (Washington, D.C.: American Association of School Administrators, 1962), p. 9.
- Klein, Stephen. "Evaluating Tests in Terms of the Information They Provide." ERIC Document ED 045 699 (June 1970).
- Kuslan, Louis, and A. H. Stone. Teaching Children Science: An Inquiry Approach. (Belmont, California: Wadsworth Press, 1968), p. 228.
- Lee, Eugene. New Developments in Science Teaching. (Belmont, California: Wadsworth Press, 1967), p. 69.
- Lindquist, E. F.(ed.) Educational Measurement. (American Council on Education, Washington, D.C., 1951), pp. 119-495.
- National Education Association. The Central Purpose of American Education. (Washington, D.C., 1961), p. 19.
- Nelson, Clarence H. "Review - Science Research Associates Achievement Series: blue version," in Buros Seventh Mental Measurements Yearbook, 2 (Highland Park, New Jersey: Gryphon Press, 1972) pp. 1231-33.
- Stake, Robert E. and T. Denny. "Needed Concepts and Techniques for Utilizing More Fully the Potential of Evaluation." The Sixty-eighth Yearbook of the National Society for the Study of Education, 2. (Chicago: University of Chicago Press, 1969).
- Thorndike, R. L. and E. Hagen. Measurement and Evaluation in Psychology and Education. (New York: John Wiley, 1955).
- Travers, Robert M. W. How to Make Achievement Tests. (New York: Odyssey Press, 1950), p. 25.

Tyler, R. W. Basic Principles of Curriculum and Instruction. (Chicago: University of Chicago Press, 1950).

_____. (ed.) Educational Evaluation: New Roles, New Means. The Sixty-eighth Yearbook of the National Society for the Study of Education, 2. (Chicago: University of Chicago Press, 1969).

Periodicals

Adams, J. F. "Test Item Difficulty and the Reliability of Item Analysis Methods." Journal of Psychology, 49, 255-262 (1960).

Anderson, Richard C. "How to Construct Achievement Tests to Assess Comprehension." Review of Educational Research, 42, 145-170 (1972).

Beard, Jean. "The Development of Group Achievement Tests for Two Basic Processes of AAAS Science - A Process Approach." Journal of Research in Science Teaching, 8, 179-183 (1971).

Billings, G. "Cognitive Levels of Elementary Science Tests." School Science and Mathematics, 71, 824-830 (December 1971).

Borton, Terry. "What's Left When School's Forgotten?" Saturday Review, 53, 69-71, 79 (April 18, 1970).

Brenner, Marshall H. "Test Difficulty, Reliability, and Discrimination as Functions of Item Difficulty Order." Journal of Applied Psychology, 48, 98-100 (April 1964).

Cole, Henry P. "Process Curricula and Creativity Development." Journal of Creative Behavior, 3, 253 (Fall 1969).

Cooley, William W. and L. E. Klopfer. "The Evaluation of Specific Educational Innovations." Journal of Research and Science Teaching, 1, 73-80 (1963).

Costin, Frank. "Optimal Number of Alternatives in Multiple-Choice Achievement Tests: Some Empirical Evidence for a Mathematical Proof." Educational and Psychological Measurements, 30, 353-358 (Summer 1970).

Davis, Fredrick B. "Item Analysis in Relation to Educational and Psychological Testing." Psychological Bulletin, 49, 97-121 (1952).

Ebel, Robert L. "Expected Reliability as a Function of Choices Per Item." Educational and Psychological Measurement, 29, 565-570 (1969).

_____. "Must All Tests Be Valid?" American Psychologist, 16, 640-647 (October 1961).

- Engelhart, Max D. "A Comparison of Several Item Discrimination Indices." Journal of Educational Measurement, 2, 69-76 (June 1965).
- Ennis, Robert H. "Needed: Research in Critical Thinking." Educational Leadership, 21, 17-20, 39 (October 1963).
- Evans, Franklin R. and R. R. Reilly. "A Study of Speededness as a Source of Test Bias." Journal of Educational Measurement, 9, 123-131 (Summer 1972).
- Feld, L. S. "Note on Use of Extreme Criterion Groups in Item Discrimination Analysis." Psychometrika, 28, 97-104 (1963).
- Flaughner, Ronald L., R. S. Melton, and C. T. Myers. "Item Rearrangement Under Typical Test Conditions." Educational and Psychological Measurement, 28, 813-824 (Autumn 1968).
- Grobman, Hulda. "Curriculum Development and Evaluation." Journal of Educational Research, 64, 436-442 (July 1971).
- Kelly, Truman. "The Selection of Upper and Lower Groups for the Validation of Test Items." Journal of Educational Psychology, 30, 17-24 (1939).
- Klosner, Naomi C. and E. K. Gellman. "The Effect of Item Arrangement on Classroom Test Performance: Implications for Content Validity." Educational and Psychological Measurement, 33, 413-418 (1973).
- Lisonbee, L. "Testing, What For?" Science Teacher, 33, 27-29 (May 1966).
- Marso, Ronald N. "Test Item Arrangement, Testing Time, and Performance." Journal of Educational Measurement, 7, 113-118 (Summer 1970).
- Morgan, D. A. "STEPS Science Test for Evaluation of Process Skills." The Science Teacher, 38, 77-79 (November 1971).
- Munz, David C. and A. D. Smouse. "Interaction Effects of Item-Difficulty Sequence and Achievement-Anxiety Reaction on Academic Performance." Journal of Educational Psychology, 59, 37-374 (October 1968).
- Nelson, Miles A. and E. C. Abraham. "Inquiry Skill Measures." Journal of Research in Science Teaching, 10, 291-297 (1973).
- Sassenrath, Julius M. "The Factorial Composition of the Iowa Tests of Educational Development." California Journal of Educational Research, 16, 80-84 (March 1965).
- Smith, Richard B. "Approach to Measurement in the New Science Curriculum." Science Education, 53, 411-415 (December 1969).
- Symonds, P. M. "Factors Influencing Test Reliability." Journal of Educational Psychology, 19, 73-87 (1938).

- Tannenbaum, R. S. "Development of the Test of Science Processes." Journal of Research in Science Teaching, 8, 123-136 (1971).
- Terranova, C. "Relationship Between Test Scores and Test Time." Journal of Experimental Education, 40, 81-83 (Spring 1972).
- Traut, Ross E. and R. K. Hambleton. "The Effect of Scoring Instructions and Degree of Speededness on the Validity and Reliability of Multiple-Choice Tests." Educational and Psychological Measurement, 32, 737-758 (1972).
- Tversky, A. "On the Optimal Number of Alternatives at a Choice Point." Journal of Mathematical Psychology, 1, 386-391 (1964).
- Tyler, Ralph W. "Resources, Models, and Theory in the Improvement of Research in Science Education." Journal of Research in Science Teaching, 5, 43 (1967).
- Walbesser, H. "Science Curriculum Evaluation: Observations on a Position," The Science Teacher, 33, 34-39 (1966).
- Welch, Wayne W. and M. O. Pella. "The Development of an Instrument for Inventorying Knowledge of the Processes of Science." Journal of Research in Science Teaching, 5, 64-68 (1967).
- Wofford, J. C. and T. L. Willoughby. "The Effects of Test Construction Variables Upon Test Reliability and Validity." California Journal of Educational Research, 20, 96-106 (May 1969).

Unpublished Works

- Fyffe, Darrel W., The Development of Test Items for the Integrated Science Processes: Formulating Hypotheses and Defining Operationally. Unpublished Doctoral Dissertation (Michigan State University, 1971).
- Robison, Richard Wayne, The Development of Items Which Assess the Processes of Controlling Variables and Interpreting Data. Unpublished Doctoral Dissertation (Michigan State University, 1973).

APPENDICES

APPENDIX I - A

ONE INDIVIDUAL COMPETENCY MEASURE FROM SAPA

Science - A Process Approach/Part G-a
Defining Operationally 7: Two Common Gases

TASK 1 (OBJECTIVE 1):

Say, If a piece of wet blue litmus paper is put into a vial of carbon dioxide gas, the color of the paper changes from blue to red. Use this information to tell me an operational definition of carbon dioxide.

Acceptable Behavior: The child states that carbon dioxide is a gas that turns wet blue litmus paper red.

TASK 2 (OBJECTIVE 1):

Say, If a piece of wet red litmus paper is put into a vial of ammonia gas, the color of the paper changes from red to blue. Use this information to tell me an operational definition of ammonia.

Acceptable Behavior: The child states that ammonia is a gas that turns wet red litmus paper blue.

TASK 3 (OBJECTIVE 2):

Show the child two vials that you have prepared previously, labeled A and B. Vial A should contain carbon dioxide and vial B ammonia gas (put two drops of clear household ammonia into the vial and cap it immediately.) Give the child two strips of red litmus paper and two strips of blue, and a vial of water. Say, One of these vials contains carbon dioxide and the other contains ammonia. Use the operational definitions of carbon dioxide and ammonia that you just stated and the objects I have given you to test the gases in the vials. Tell me which vial contains carbon dioxide and which contains ammonia.

Acceptable Behavior: The child moistens the strips of litmus paper, uncaps one vial, puts one strip of red and one strip of blue litmus paper into it, quickly caps the vial, and observes the papers to see which changes color. He does the same with the other vial. He states that he concludes from his observations that vial A contains carbon dioxide and vial B contains ammonia.

TASK 4 (OBJECTIVE 3):

Prepare some ammonia gas by placing a piece of paper toweling in the bottom of a 50-milliliter vial, and adding about 1 milliliter of clear household ammonia to it. Invert another 50-milliliter vial over the first, let it stand for about a minute, and then remove and quickly cap it. (See Figure C). Show the child the capped vial. It will

Appendix I-A cont'd

contain mostly air, but enough ammonia gas to give a satisfactory test for this task. Also show him some bromthymol blue solution. Say, Watch while I add some of this green test liquid to this vial of ammonia gas. Draw about 1 milliliter of the liquid into a medicine dropper. Uncap the vial of ammonia, squirt the green test liquid into the vial and quickly replace the cap. The green test liquid will immediately turn bright blue. Say, One operational definition of ammonia is that it is a gas that turns moist red litmus paper blue. Tell me an alternate operational definition of ammonia based on your observations of what happened with green test liquid.

Acceptable Behavior: The child states that ammonia is a gas that turns green test liquid blue.

APPENDIX I - B

INTEGRATED PROCESSES OF SAPA

INTERPRETING DATA

Under the general heading of Interpreting Data the following skills are stressed.

- "1. DESCRIBE in a few sentences the information 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 of linear relations, and DESCRIBE the information provided by the slope."¹

CONTROLLING VARIABLES

Under the general heading of Controlling Variables, Science - A Process Approach attempts to develop in the child the following skills in working with variables.

- "1. IDENTIFY variables which may influence the behavior or the properties of a physical or biological system.
2. IDENTIFY variables which are held constant, manipulated, or responding 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."²

¹American Association for the Advancement of Science, Science - A Process Approach Commentary for Teachers, AAAS Misc. Publication 68-7 (1968), p. 187.

²Ibid, p. 177.

Appendix I-B cont'd

FORMULATING HYPOTHESES

Under the general heading of Formulating Hypotheses the following skills are stressed:

- "1. CONSTRUCT a hypothesis that is a generalization of observations or that is a generalized explanation.
2. CONSTRUCT and DEMONSTRATE a test of a hypothesis.
3. DISTINGUISH between observations that support a hypothesis and those that do not.
4. CONSTRUCT a revision of a hypothesis based on observations that were made to test the hypothesis."³

DEFINING OPERATIONALLY

"In defining operationally physical scientists state 'What you do or what operation you perform' and 'what you observe.' For example applying these criteria an operational definition of oxygen might be:

Oxygen: A gas that causes a glowing splint to burst into flame (what you observe), when the splint is placed (what you do) into a container of the gas.

If a child wishes to decide if a gas is oxygen using this definition he knows exactly what to do and to observe. In contrast a non-operational definition of oxygen, as far as a child is concerned, would be: Oxygen is an element composed of atoms having atomic number 8 and atomic weight 16. Given a container of gas this definition will be entirely useless to the child. He will know neither what to do nor what to observe."⁴

³Ibid, p. 159.

⁴Ibid, p. 167.

APPENDIX I - C

LISTING OF INDIVIDUAL COMPETENCY MEASURES

Individual Competency Measures used in this study are indicated by an asterisk.

			<u>Number of Tasks</u>
Interpreting Data:			
* E/d	ID 1	Guinea Pigs in a Maze	9
* E/l	ID 2	Identifying Materials	6
* E/o	ID 3	Precision in Measurement	4
* E/u	ID 4	Field of Vision	3
F/c	ID 5	Magnetic Fields	3
F/g	ID 6	Quantitative Analysis	10
F/l	ID 7	A Measure of Chance	8
F/p	ID 8	Contour Maps	6
F/q	ID 9	Measuring Small Things	4
G/e	ID 10	Moon Photos	5
Total Tasks Used			22

Controlling Variables:

* E/b	CV 1	Rolling Cylinders	10
E/c	CV 2	Upward Movement of Liquids	4
E/p	CV 3	Growth of Mold on Bread	5
* E/q	CV 4	Loss of Moisture from Potatoes	7
* F/a	CV 5	Variables Affecting Chemical Reactions	10
* F/e	CV 6	The Effects of Practice on Memorization	15
* F/h	CV 7	Nutrition of a Small Animal	8
* F/j	CV 8	Forgetting and Relearning	4
F/m	CV 9	Human Reaction Time	8
F/o	CV 10	Growth and Orientation of Plants	5
F/r	CV 11	A Small Water Animal	5
G/e	CV 12	Precipitating Salts from Solution	4
Total Tasks Used			54

Formulating Hypotheses:

* E/h	FH 1	Observations and Hypotheses	4
* E/i	FH 2	Conductors and Nonconductors	4
* F/b	FH 3	Effects of Temperature on Dissolving Time	6
* F/i	FH 4	Levers	3
* F/k	FH 5	Tasters and Nontasters	6
* G/d	FH 6	Variation in Perceptual Judgment	6

Total Tasks Used 29

Appendix I-C cont'd

Defining Operationally:			<u>Number of Items</u>
* E/i	DO 1	Electric Circuits and their Parts	3
* E/m	DO 2	Analysis of Mixtures	3
E/v	DO 3	Living Things are Composed of Cells	6
* F/d	DO 4	Determining the Direction of True North	3
* F/f	DO 5	Inertia and Mass	4
F/n	DO 6	Parts of Living Plants	3
G/a	DO 7	Two Common Gases	4
* G/b	DO 8	Temperature and Heat	3
Total Tasks Used			16

APPENDIX III-A

THE MINIMUM LEVEL OF DISCRIMINATION - CONVENTIONAL ITEM ANALYSIS

For the purpose of observing what effect changing the minimum acceptable level of discrimination has on TSPT form C, the following experiment was conducted: Using data obtained from the validation sample and conventional procedures for calculating the discrimination index, revisions of form C were performed by computer selection of items based solely on minimum acceptable discrimination index. The results are displayed below:

N=52

<u>Item Selection</u> <u>Criterion</u>	<u>K</u> <u>Items</u>	<u>Percent</u>		<u>KR20</u>	<u>ICM</u> <u>Corr.*</u>
		<u>Diff.</u>	<u>Disc.</u>		
TSPT form C	61	56	38	0.86	0.78
Min. Disc. 0.2	48	53	48	0.89	0.81
0.3	36	52	56	0.90	0.81
0.4	30	50	60	0.88	0.82
0.5	22	48	65	0.88	0.83

*Correlation with the Individual Competency Measures Scores.

These data reveal that a fairly low level requirement for the minimum discrimination produces a noticeable improvement in most of the statistics; but, probably as a result of loss in number of items, as the minimum discrimination requirement is further increased, the KR20 value begins to drop. Thus, for this study, probably the 0.3 minimum discrimination requirement used in the item improvement phase seems reasonable.

APPENDIX III-B
DIRECTIONS FOR ADMINISTERING TSPT FORM D

APPENDIX III-B

DIRECTIONS FOR ADMINISTERING THE SCIENCE PROCESSES TEST

(TSPT)

TIME REQUIRED:

TSPT is not intended to be a timed test. Thus you need not be concerned that all the students start and stop at exactly the same instant. Students should be encouraged to work efficiently but should not feel pressured by time limitations. Most students will complete TSPT in about 45 minutes.

MATERIALS:

The only materials the student will need in addition to the TSPT test booklet and the TSPT answer sheet are a pencil (number 2 is recommended) and an eraser.

MARKING THE ANSWER SHEET:

Since the answer sheets will be machine scored, the students response marks need to be dense and black, and should approximately fill the response box without extending beyond it. Mistakes should be erased cleanly. Although students should be encouraged to use reasonable care in marking. Extreme concern on this point is not necessary.

USING THE NAME BLOCK:

The name block need not be filled in if you anticipate that the students will have difficulty with it. The students need only turn the answer sheet side-wise and print their last name and first name in the boxes provided at the top of the name block, placing one letter in each box. If a student's name is too long to fit in the boxes, tell him to simply leave off the last few letters of his name.

DIRECTIONS TO THE STUDENTS:

The directions to be read to the students are set off by vertical lines. These need not be read word for word, but may be paraphrased or amplified as desired.

Opening Statement:

Many experts feel that the standardized tests you take every year (SRA, ITED, etc.) are not quite fair to you because they ask you about a lot of facts while the science you study in school tries to teach you more about how to think and how to act like a scientist. Today you are being given a chance to help in making a new test - TSPT which will measure your knowledge of the processes of science. If you work hard and do your best on this test, we should be able to tell you how well you are learning the processes of science and you will be helping us to make a more fair test. There are 36 questions on this test. If you work at a steady pace you should have plenty of time to finish. The answer sheet will be handed to you first. Do not write on it until you are told what to do. You will need a pencil with a good eraser for marking the answer sheet.

Appendix III-B cont'd

Distribute the answer sheets; one to each student. Check to see that each student has a pencil and an eraser.

Using the Answer Sheet:

Turn the answer sheet side-wise so the letters TSPT are at the bottom below the name block. (Hold up an answer sheet turned correctly.) Print your last name and your first name in the boxes at the top of the name block. Put one letter in each box. If your name is too long to fit in the boxes, leave off the last letters.

If your students are familiar with the use of the name block, you may instruct them to fill it in. Otherwise, tell them to ignore it.

You can see that there is one line on the answer sheet for each page of the test. For example, page 1 has only question 1 on it, while page 2 has questions 2, 3, 4, and 5 on it, etc. You are to black in the little box just below the letter which you feel is the best answer for each question. Be sure to mark only one answer for each question. If you make a mistake, erase the mistake. Since your answer sheet will be read by a machine, erase cleanly and make your marks dark. Fill, but do not go outside the little box. Since others will be using the test booklets, do not make any marks on them. As soon as you are given the test booklet, you may open it and begin work.

Distribute the test booklets. Check to see that the students have entered their names correctly in the name block.

DURING THE TESTING PERIOD:

Check to see that the students are marking the answer sheet properly.

FOLLOWING THE TEST:

Collect the test booklets and answer sheets. Place them in the container provided and return it to the principal's office.

Thank you for your help.

APPENDIX III-C
TSPT FORM D TEST MANUAL

TEST MANUAL

THE
SCIENCE
PROCESSES
TEST

FORM D
(1974)

by

ROBERT R. LUDEMAN
DARRELL W. FYFFE
RICHARD W. ROBISON
RICHARD J. MCLEOD
GLENN D. BERKHEIMER

COPYRIGHT BY
ROBERT R. LUDEMAN
1974

USE THE ANSWER SHEET PROVIDED
PLEASE DO NOT MAKE ANY MARKS ON THIS BOOKLET

THE SCIENCE PROCESSES TEST
(TSPT)

TEST MANUAL

by

Robert R. Ludeman
Andrews University
Berrien Springs, Michigan

Darrel W. Fyffe
Bowling Green State University
Bowling Green, Ohio

Richard W. Robison
Manchester College
North Manchester, Indiana

Richard J. McLeod
Glenn D. Berkheimer
Michigan State University
East Lansing, Michigan

Rationale

A more complete description of the rationale and development of TSPT is published elsewhere.¹

The expressions of need voiced by researchers in science education for efficient valid tests of science processes coupled with the difficulty usually encountered in developing such tests prompted a group of individuals in the Science and Math Teaching Center at Michigan State University to employ a method of test development that is different from the traditional test development procedure. TSPT is the result of that effort.

¹Robert R. Ludeman, Development of The Science Processes Test, unpublished dissertation (Michigan State University, 1974).

Most tests rely on "expert" opinion for their claim of content type validity. This procedure is especially subject to question for a test intended to evaluate children's ability to use the processes of science since it has been found that writing process test items is much more difficult than writing simple factual recall items. Therefore, in the development of TSPT, although this procedure was used for the original generation of test items, in the later stages of test development, this procedure was replaced by a procedure known as "external criterion referenced validation." Using this procedure, items are included in the test on the basis of the requirement that children's performance on each item be highly correlated with their performance on the external criterion. In this case, the external criterion is the Individual Competency Measures of the elementary science program Science - A Process Approach (SAPA).²

The Individual Competency Measures consist of individualized testing situations using the same materials and contexts used by SAPA in defining the Science Processes. The test administrator evaluates the subject's ability to use the Processes as he works with materials in solving problems the administrator poses. The Individual Competency Measures are not widely used because of their low time-efficiency but since they are so directly related to the context which defines the Processes, it seems reasonable to assume that they constitute an accurate assessment of student's ability and can be used as the criterion for validation of a more time-efficient test.

The science processes addressed by TSPT are the integrated processes referred to by SAPA as Interpreting Data, Controlling Variables, Formulating Hypotheses, and Defining Operationally. It is assumed that the Individual Competency Measures do indeed measure children's ability

²American Association for the Advancement of Science, Science - A Process Approach Commentary for Teachers, AAAS Misc. Publication 68-7, 1968.

to use these processes and that a high correlation of children's performance on the Individual Competency Measures with their performance on TSPT therefore may be taken as evidence that TSPT is a valid measure of children's ability to use the processes of science.

Development of TSPT:

Based on the behavioral objectives of SAPA, 113 multiple choice items were originally written and examined by science educators with reference to their relevance to these objectives. These items were tried out and revised three times on the basis of conventional test development procedures (all alternatives chosen by some students, discrimination greater than .3, difficulty between .2 and .7) with many items being either discarded or rewritten to meet these requirements. In this phase of the development which was completed in the spring of 1973, 367 sixth-grade students were involved. At this point it was felt that the resulting item pool of 61 items was of adequate technical quality to begin the criterion-validation phase of the development. Accordingly, beginning early in November and continuing through December of 1973 the Individual Competency Measures of SAPA were administered to 52 sixth-grade children. Immediately on completion of the administration of the Individual Competency Measures the above 61 items were administered to the same 52 children. Their performance on the Individual Competency Measures was then used as the criterion for item selection for inclusion in TSPT, using the following requirements:

1. All alternatives have been chosen by some students.
2. The context of the item allows its use.

In some cases, since more than one item was based on a given context, the group of items had to be included or excluded in toto.

3. The difficulty of each item (proportion of students missing the item) was required to be between .2 and .7.

4. Using the Individual Competency Measures scores to define the "upper 27 percent" and "lower 27 percent" groups, each item was required to have a minimum discrimination of .2.
5. The correlation of students' scores on each item with their scores on the Individual Competency Measures was required to be .2 or greater.

Out of the above items which met these requirements, 36 were used to make up TSPT. Although more items might have been included and would have been desirable from a strictly statistical viewpoint, experience gained during item try-outs indicated that if the number of items exceeded about 40, the students began to get restless and lose their concentration before they finished the test. TSPT was then printed and a machine scoreable answer sheet was designed and printed.

A summary of various correlations obtained from the above procedure is listed in Table 1.

Table 1 - Correlation Summary
N = 52

TSPT - ICM*	.830
TSPT - SRA** Science	.788
TSPT - SRA** Reading	.798

*Individual Competency Measures

**Science Research Associates Achievement
Series (blue version)

Norming TSPT:

A Norming Sample was selected from the public schools containing sixth-grade classes as listed in the Michigan and Indiana public school directories and which are located within a 50 mile radius of Andrews

University in Berrien Springs, Michigan. From this population of 243 schools a random sample of 20 schools was drawn. One of these schools refused to participate in the study so the actual norming sample consisted of 19 schools from 12 different school systems. The sample contained rural, suburban and city schools in about equal numbers. The largest school contained 168 sixth grade students and the smallest contained 21 sixth grade students. There was a total of 1301 students in the norming sample, with a broad spectrum of science programs represented. Since no systematic relation was observed between students' scores and type of science program studied, no effort is made to distinguish among programs used by the norming sample. TSPT form D was administered to this norming sample by their own teachers in their own classrooms in the spring of 1974. The important test statistics obtained from the norming sample is displayed in Table 2. The distribution of student scores is given in Table 3.

Table 2 - NORMING SAMPLE DATA FOR TSPT

Grade level	6
Number of Items	36
Number of Subjects	1301
Median Score	17
Mean Score	17.9
Standard Deviation	6.90
Standard Error of the Measurement	2.69
Mean Point Biserial Correlation	.409
KR20 Reliability	.842
Mean Difficulty	.503
Mean Discrimination	.496

Table 3 - Norming Sample Distribution
N = 1301

<u>Raw Score</u>	<u>Frequency</u>	<u>Std. Score</u>	<u>Percentile</u>
36	0	+2.62	
35	0	2.48	
34	3	2.33	99.8
33	6	2.19	99.3
32	13	2.04	98.3
31	21	1.90	96.7
30	24	1.75	94.9
29	27	1.61	92.8
28	31	1.46	90.4
27	47	1.32	86.8
26	47	1.17	83.2
25	50	1.03	79.3
24	45	0.88	75.9
23	58	0.74	71.4
22	61	0.59	66.7
21	53	0.45	62.6
20	48	0.30	59.0
19	51	0.16	55.0
18	63	+0.01	50.2
17	59	-0.13	45.7
16	55	0.28	41.4
15	55	0.42	37.2
14	72	0.57	31.7
13	69	0.71	26.4
12	65	0.86	21.4
11	70	1.00	16.0
10	59	1.15	11.5
9	52	1.29	7.5
8	34	1.43	4.8
7	32	1.58	2.4
6	16	1.72	1.2
5	7	1.87	0.6
4	6	2.01	0.2
3	1	2.16	0.1
2	1	2.30	0.0
1	0	2.45	
0	0	-2.59	

TSPT is intended to be a "power test" so ample time should be given for essentially all students to complete the test. For the norming sample it was found that 45 minutes was adequate.

Reading Level:

Attention was given during item writing and editing to keeping the reading level as low as possible. The resulting reading level for the final test using the reading scale developed by Fry³ is approximately low sixth-grade.

In instances where classes were segregated on the basis of "good readers" and "slow readers" the "good readers" typically scored about 5 points higher than the "slow readers."

DIRECTIONS FOR ADMINISTERING TSPT:

TIME REQUIRED:

At least 45 minutes without interruption should be provided for the administration of TSPT.

TSPT is not intended to be a timed test. Thus you need not be concerned that all the students start and stop at exactly the same instant. Students should be encouraged to work efficiently but to take time to think through their answers. TSPT is not a factual recall test. Thinking is required to achieve a high score on this test. Most students will complete TSPT in less than 45 minutes.

MATERIALS:

The only materials the student will need in addition to the TSPT test booklet and the TSPT answer sheet are a pencil (number 2 is recommended) and an eraser.

MARKING THE ANSWER SHEET:

Since the answer sheets are intended to be machine scored, the student's response marks should be distinct and should approximately

³Edward B. Fry, Reading Instruction for Classroom and Clinic, (McGraw Hill, New York, 1972).

fill the response box without extending beyond it. A single dark mark is preferred. Mistakes should be erased cleanly. Although students should be encouraged to use reasonable care in marking, extreme concern on this point is not necessary.

USING THE NAME BLOCK:

First the student should turn the answer sheet sidewise and print the letters for his name in the boxes provided at the top of the name block, one letter in each box. Care must be taken that the first letter of the last name is entered in the first box. If a student's name is too long to fit in the boxes provided, the last few letters should be omitted. In order for the machine to read the name, the letter in each alphabet column in the name block corresponding to the letter the student has placed in the box at the top of each column of the name block must be marked in. A single clean mark which approximately fills but does not extend beyond the response box is required. Only one letter may be marked in each column of the name block.

DIRECTIONS TO THE STUDENTS:

The directions to be given to the students are set off by vertical lines. These need not be read word for word, but may be paraphrased or amplified as desired.

Opening Statement:

This test will find out how well you can use the processes of science. That is, how well you can think and answer the way a scientist would. This means you will need to take time to think before you can answer the questions. You will have enough time so do not rush. If you work at a steady pace, you will have plenty of time to finish. The answer sheet will be handed to you first. Do not write on

it until you are told what to do. You will need a pencil with a good eraser for marking the answer sheet.

Distribute the answer sheets, one to each student. Check to see that each student has a pencil and an eraser.

Using the Answer Sheet:

Turn the answer sheet sideways so the letters TSPT are at the bottom below the name block.

Hold up an answer sheet turned correctly.

Print your last name and your first name in the boxes at the top of the name block. Be sure you begin with the first box and put one letter in each box. If your name is too long to fit in the boxes, just leave off the last few letters.

Allow sufficient time for the names to be entered. Spot-check to see that it is done correctly.

Under the box where you printed the first letter of your name, go down the alphabet until you come to the first letter of your name. Draw a line through that letter. Be careful that your mark does not go outside the little box. Under the box where you printed the second letter of your name, go down that alphabet until you come to the second letter of your name. Draw a line through that letter. Do this for all the rest of the letters of your name.

Allow sufficient time for the name block to be filled. Spot-check to see that this is done correctly.

Turn the answer sheet right-side-up. You can see that there is one line on the answer sheet for each page of the test. For example, page one has only question one on it, while

page 2 has questions 2, 3, 4, and 5 on it, etc. You are to draw a line through the little box just below the letter which you feel is the one best answer for each question. If you make a mistake, erase it cleanly. Make your marks go the whole length of the little boxes, but be sure they do not go outside the little boxes. Since others will be using the test booklets, do not make any marks on them. As soon as you are given the test booklet, you may open it and go to work.

Distribute the test booklets. At the same time, check to see that the students have filled out the name block correctly.

DURING THE TESTING PERIOD:

Check to see that the students are marking the answer sheets properly.

FOLLOWING THE TEST:

Separate the test booklets and the answer sheets. Arrange the answer sheets in the order in which you wish the result returned to you.

Return all test booklets and answer sheets to R. Ludeman, Andrews University, Berrien Springs, Michigan 49104. The answer sheets will be machine scored and returned to you together with a computer printout of student's scores and test statistics similar to what appears in Table 2 of this manual. By special arrangement, other data may be obtained as well, such as item analysis information, breaking the test down into subtests, correlation of two sets of scores, etc.

INTERPRETING TSPT SCORES

Care should be exercised in using both Tables 2 and 3 for interpreting the results of any administration of TSPT. The norming sample used to obtain these data should not be assumed to be representative of any wider population than that previously described.



APPENDIX IV-A

TSPT FORM A

THE SCIENCE PROCESSES TEST

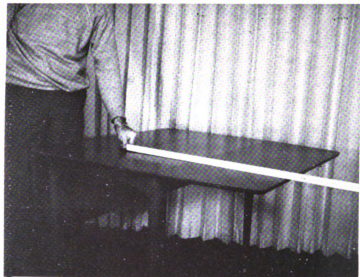
MICHIGAN STATE UNIVERSITY

Use the answer sheet provided.

Please do not make any marks on this test booklet.

1. A stick 100 centimeters long is slowly pushed over the edge of a table as shown. About how far over the edge do you think the end will reach before the stick will fall?

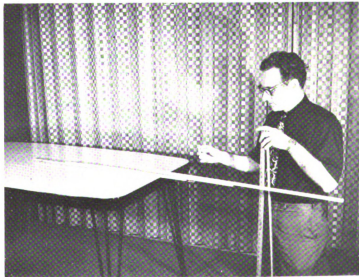
- a. 10 cm.
- b. 30 cm.
- c. 50 cm.
- d. 70 cm.
- e. 90 cm.



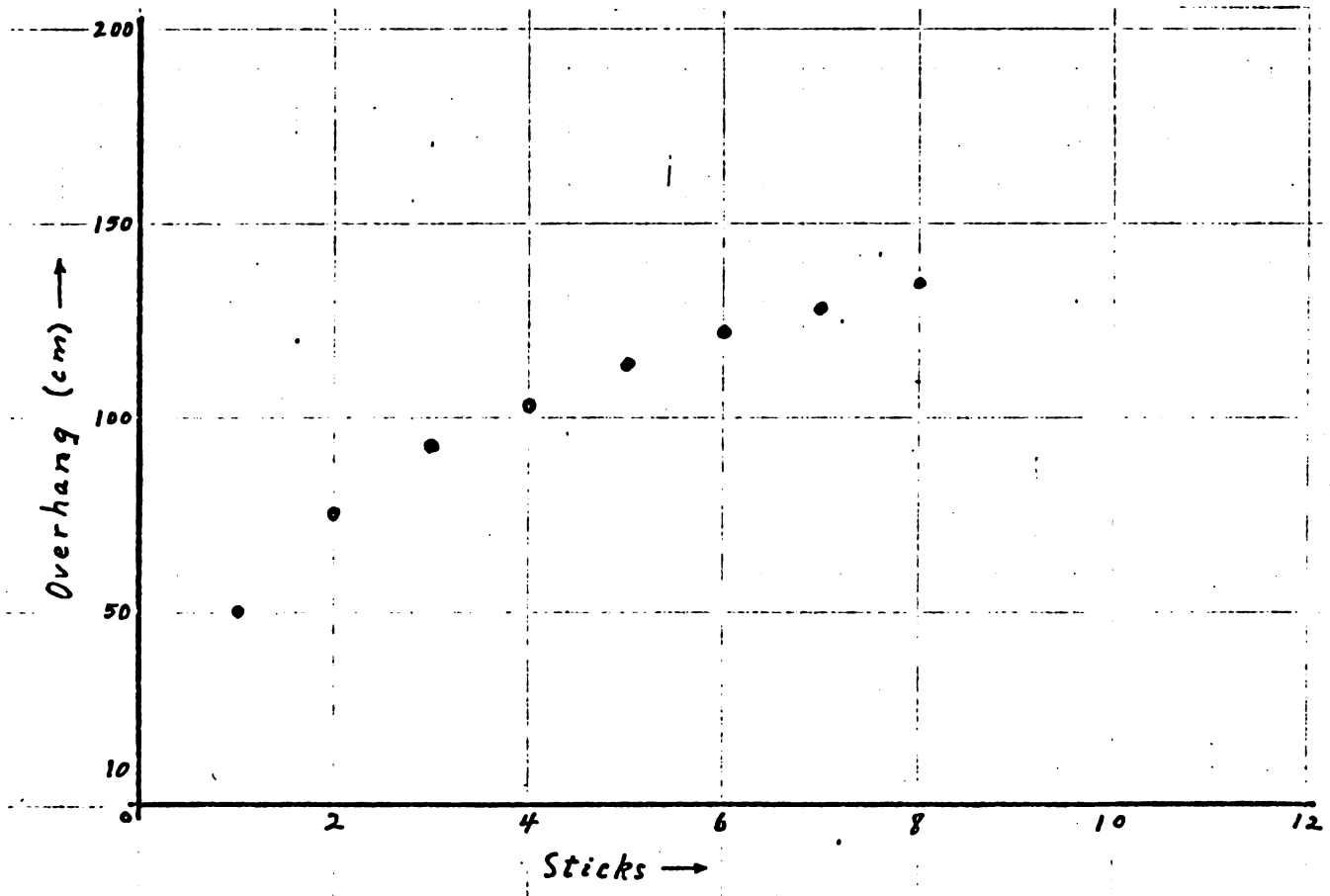
2. The picture that would show the answer to the above question best would be a picture showing
- a. The stick balanced on my finger.
 - b. How thick the stick is.
 - c. The stick after it has fallen off the table.
 - d. Any of the above pictures would give the answer.
 - e. None of the above pictures would give the answer.

Questions 3, 4 and 5 use the following set up:

Several identical sticks are stacked on top of each other and extended over the edge of a table in such a way as to give the greatest possible overhang (see below).



The relation between maximum overhang and number of sticks is graphed below:

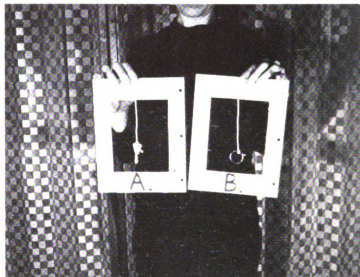


3. The greatest overhang you could get using 5 sticks would be about:
- a. 1 cm.
 - b. 49.9 cm.
 - c. 99 cm.
 - d. 112 cm.
 - e. None of the above is close.
4. The smallest number of sticks you would need to get an overhang of 100 cm. is:
- a. 1 stick
 - b. 2 sticks
 - c. 3 sticks
 - d. 4 sticks
 - e. you could not ever get that big an overhang.
5. Using 10 sticks you could get a maximum overhang of:
- a. more than 150 cm.
 - b. between 140 and 150 cm.
 - c. between 130 and 140 cm.
 - d. less than 130 cm.
 - e. there is not enough data to decide which is the best answer.

6. If you needed to tell someone what I mean above by "overhang" so that they could measure the overhang, you should say "overhang" means the distance from the:

- a. end of the top stick to the center of gravity of the system.
- b. end of the top stick to the edge of the table.
- c. center of gravity to the edge of the table.
- d. more than one of the above is correct.
- e. none of the above is correct.

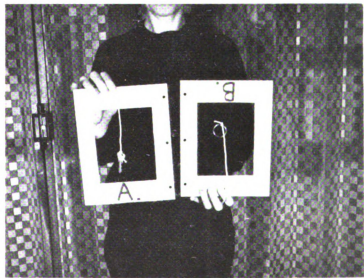
Questions 7 through 10 are about frames A and B shown in the picture on the right.



In this picture frame B has been turned upside down.

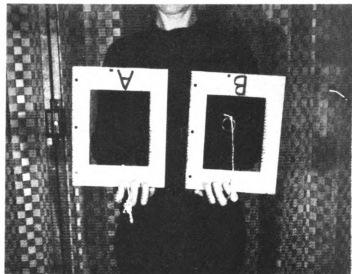
7. Based on what you have seen this far it is possible that:

- a. string A will remain straight when frame A is turned upside down.
- b. string A will bend when frame A is turned upside down.
- c. string B is held straight by a fine thread fastened to the bottom of the frame.
- d. all of the above may be true.
- e. none of the above can possibly be true.



8. From the picture on the right you now know that string:

- a. A is not as stiff as string B.
- b. A is made of a stiff wire that is now bent.
- c. B is made of a stiff wire.
- d. B is held up by a strong magnet hidden behind the frame.
- e. none of the above is correct.

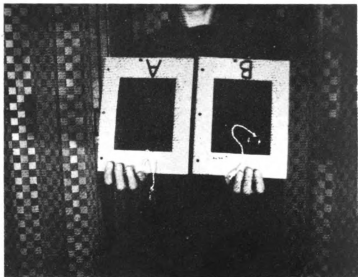


9. What evidence do you now have that something about frame A is different different from frame B.

- a. the key in A fell.
- b. the ring in B did not fall.
- c. either of the above is evidence there is a difference.
- d. Both 1 and 2 are needed to have evidence for a difference.
- e. none of the above is correct.

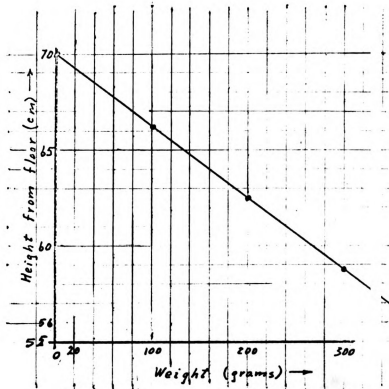
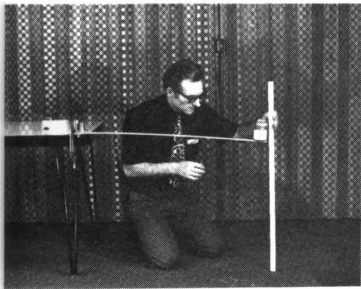
10. Based on the evidence you now have (including the picture on the right), the best conclusion is:

- a. string A is not as stiff as string B.
- b. string A is made of a stiff wire that is now bent.
- c. string B is made of a stiff wire.
- d. B is held up by a strong magnet hidden behind the frame.
- e. none of the above is reasonable.



11. My name for the special string used in frame B above is "Wyrstring." Suppose a friend phoned you to find out if a piece of string he found was wyrstring. To tell him it would be most helpful to know:

- how stiff his string is.
- how long his string is.
- how big around his string is.
- what his string is made of.
- where he found his string.

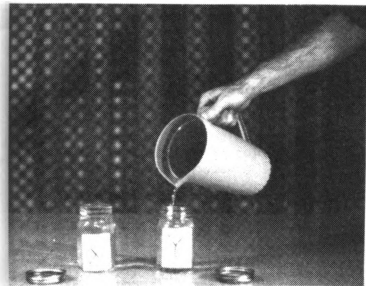


The graph on the right was obtained by setting weights on the end of the stick clamped to the table as shown on the left.

12. The can of soup shown above bends the stick to 66 cm from the floor. The can weighs about:
- 67.5 grams
 - 110 grams
 - 255 grams
 - 315 grams
 - none of the above.

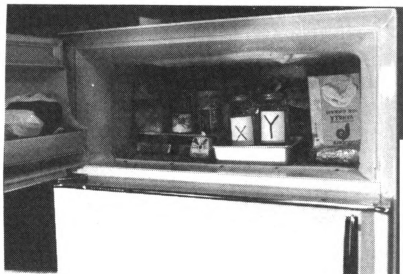
13. A weight of 65 grams should bend the stick to about
- 57.5 cm from the floor
 - 64 cm from the floor
 - 67.5 cm from the floor
 - 135 cm from the floor
 - none of the above.
14. The weight that would bend the stick to 72 cm from the floor would weigh about:
- 55 gm
 - 67.5 gm
 - 215 gm.
 - 350 gm
 - none of the above
15. According to the above experiment, doubling the weight on the end of the stick should:
- double its distance from the floor.
 - cut its distance from the floor in half.
 - bend it down about 5 cm closer to the floor.
 - bend it closer to the floor but not by a fixed distance.
 - none of the above is correct.

Questions 16 through 22 are about the following experiment



The two jars shown are filled full to the brim.

The lids are screwed on tight.



The jars are put in a pan and placed in a freezer.

The temperature is set at 0 degrees F.

16. Two hours later it is found that neither jar is frozen, possibly because:
- neither contains water.
 - more time is needed.
 - they are too full to freeze.
 - all the above are reasonable.
 - two of the above answers are reasonable.
17. One might expect both jars to be frozen solid one day later because:
- water freezes at temperatures below 32 degrees F.
 - the liquids in the jars look like water.
 - one day in a freezer should be long enough to freeze water.
 - all of the above are true.
 - none of the above is true.

Next day when the freezer is opened jar "Y" is broken and its contents frozen solid.

18. You now know that:

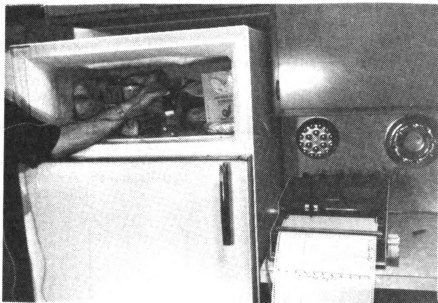
- the jars contents are different.
- at least one of the jars contains water.
- the temperature of the jars is different.
- more than one of the above is correct.
- none of the above is correct.



19. Suppose someone told you that the contents of jar "Y" behaved like a Cronon while the contents of jar "X" behaved like a non-Cronon. To show the difference between them you can say a Cronon is:
- a chemical, but a non-Cronon is not.
 - just another name for water.
 - easier to freeze than a non-Cronon.
 - more than one of the above is correct.
 - none of the above is correct.

20. A jar of water is left in the above freezer over night and is found to be frozen next morning. You know that:
- a. water freezes easier than a Cronon.
 - b. water is a Cronon.
 - c. Cronons are made of water.
 - d. the water was cold before it was put in the freezer.
 - e. none of the above is correct.
21. A bottle of alcohol is left in the above freezer over night. It is not frozen. You now know that:
- a. alcohol is a Cronon.
 - b. alcohol is a non-Cronon.
 - c. non-Cronons are made of alcohol
 - d. alcohol cannot be frozen.
 - e. none of the above is correct.
22. From this experiment it is safe to say that a mixture of half water and half alcohol would:
- a. be a Cronon.
 - b. be a non-Cronon.
 - c. freeze easier than pure alcohol.
 - d. freeze easier than pure water.
 - e. none of the above is correct.
23. Richard claims that any glass bottle will break when the water it contains freezes. To test this idea, he puts four bottles in the freezer: A is empty, B is one-third full of water, C is two-thirds full, D is brim-full. If Richard is correct, which bottles will break?
- a. A only
 - b. A, B, and C.
 - c. B and C.
 - d. B, C, and D.
 - e. all the bottles.

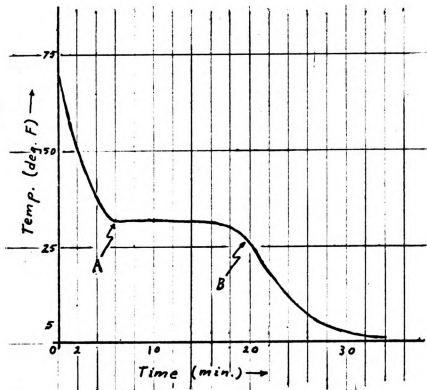
On the right is a picture of a recording thermometer. The thermometer is the dark object being placed in the beaker. It sends an electrical signal to the recorder on the left whose pen automatically draws a graph of temperature and time.



24. 100 ml of water was placed in the above beaker and left in the freezer.

The recording thermometer drew the graph on the right. Thus you know that a 10 degree drop in temperature:

- requires a time of 5 minutes.
- requires a time of 15 minutes.
- requires less time for colder temperatures.
- requires more time for colder temperatures.
- none of the above is correct.

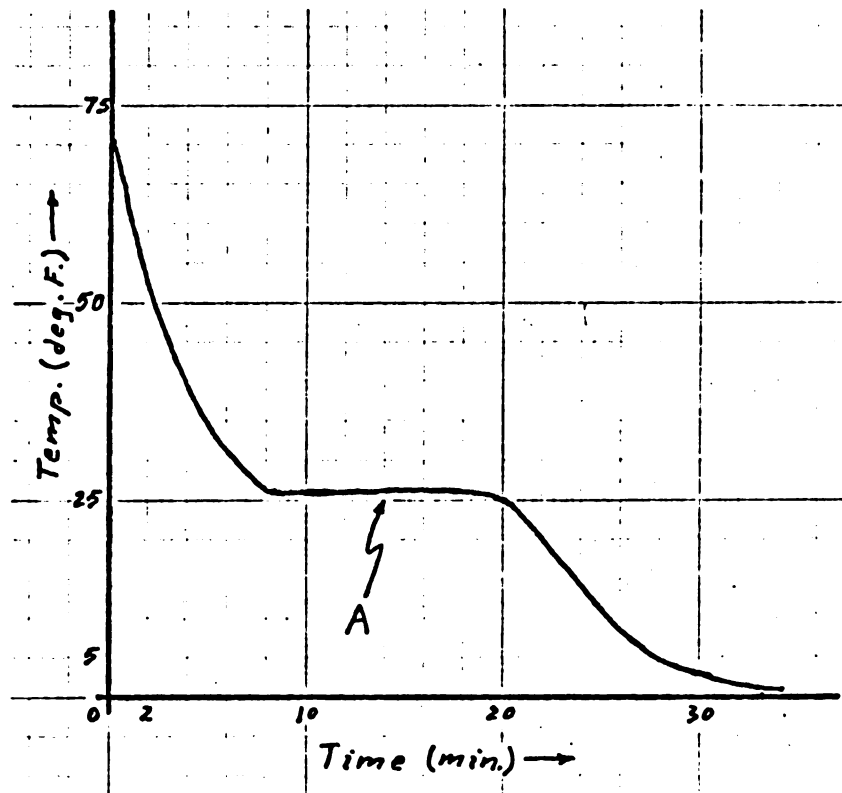


25. When the recorder reached point A above I opened the freezer door for a quick look. Ice was just beginning to form on the surface of the water. I looked again at point B. The water was all frozen solid. From what you know now it is possible that the cause of the "flat spot" on the graph is:
- the recorder sticks at about 32 degrees and so it does not read right.
 - the temperature does not change while the water is turning to ice.
 - opening the freezer door ruined the experiment.
 - more than one of the above are reasonable.
 - none of the above are reasonable.

26. Another 100 ml beaker of water has 10 grams of salt dissolved in it. It is placed in the freezer.

The recording thermometer draws the graph on the right. The salt seems to have had the greatest affect on:

- the temperature of the "flat spot."
- the amount of time for the "flat spot."
- the cooling rate before the "flat spot."
- the cooling rate after the "flat spot."
- none of the above were affected.



27. When the recorder reached point A I opened the freezer door for a quick look. About half of the salty water was frozen. From what you know now it is probable that the cause of the "flat spot" is:
- the recorder sticks at about 32 degrees and so it does not read right.
 - the temperature does not change while the liquid is turning to a solid.
 - opening the freezer door ruined the experiment.
 - more than one of the above are reasonable.
 - none of the above are reasonable.

28. For future experiments it would not be necessary to open the freezer door to determine when freezing is taking place because we can say freezing occurs:
- during the time when the cooling curve is temporarily flattened.
 - when a regular crystal structure develops.
 - at constant molecular energy.
 - when molecular motion ceases.
 - none of the above is correct.
29. Between the experiments of problems 24 and 26 I changed the following:
- the temperature of the water.
 - the temperature of the freezer.
 - the time.
 - the amount of salt in the water.
 - all of the above.
30. The following variable/s was/were kept constant in the experiment:
- the temperature of the water.
 - the temperature of the freezer.
 - the amount of salt in the water.
 - all of the above.
 - none of the above.
31. This experiment tested the idea that:
- the time for water to freeze depends on the amount of water used.
 - the time for water to freeze depends on the temperature of the freezer.
 - the temperature at which water freezes depends on the amount of water used
 - the temperature at which water freezes depends on the temperature of the freezer.
 - none of the above.

THE SCIENCE PROCESSES TEST

MICHIGAN STATE UNIVERSITY

Use the answer sheet provided.

Please do not make any marks on this test booklet.

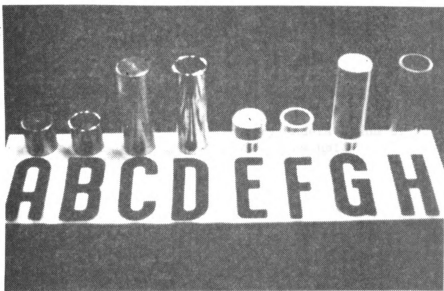
Questions 32 - 80

Questions 32 to 45 are about the cylanders shown below.

Cylinders A to D are made of metal while cylinders E to H are made of clear plastic.

Cylinders A, B, E and F are short while cylinders C, D, G, and H are long.

Cylinders A, C, E, and G are solid, while cylinders B, D, F and H are hollow.



It is expected that the time it takes these cylinders to roll the length of the sloping table on the right will depend on some or all of the above variables.

32. By comparing rolling times for cylinders A and E you could test the effect of the variable:

- a. solid or hollow.
- b. long or short.
- c. metal or plastic.
- d. amount of slope.
- e. none of the above.



33. By comparing rolling times for cylinders F and H you could test the effect of the variable:

- a. solid or hollow.
- b. long or short.
- c. metal or plastic.
- d. amount of slope.
- e. none of the above.

34. By comparing rolling times for cylanders A and G you could test the effect of the variable:
- a. solid or hollow.
 - b. long or short.
 - c. metal or plastic.
 - d. amount of slope.
 - e. none of the above.
35. It was found that the rolling time was exactly the same for cylanders B and D. From this information alone, which of the following variables does not affect the rolling time:
- a. solid or hollow.
 - b. long or short.
 - c. metal or plastic.
 - d. amount of slope.
 - e. none of the above.
36. The rolling times for which cylanders will tell you if a hollow cylander rolls at a different rate than a solid cylander?
- a. A and D.
 - b. A and F.
 - c. C and D.
 - d. C and H.
 - e. none of the above.
37. The rolling times for which cylanders will tell you if a metal cylander rolls at a different rate than a plastic cylander?
- a. A and F.
 - b. A and H.
 - c. C and G.
 - d. C and H.
 - e. none of the above.

Rolling time for the above cylanders is given in the following table. Use it to answer questions 38 to 45.

<u>Rod</u>	<u>Material</u>	<u>Length</u>	<u>Type</u>	<u>Time</u>
A	metal	2 cm	solid	5 sec.
B	metal	2	hollow	10
C	metal	8	solid	5
D	metal	8	hollow	10
E	plastic	2	solid	5
F	plastic	2	hollow	10
G	plastic	8	solid	5
H	plastic	8	hollow	10

38. The material from which the cylander is made affects the rolling time.

- a. true.
- b. false.
- c. cannot tell from the data.

39. The variable solid or hollow affects the rolling time.

- a. true.
- b. false.
- c. cannot tell from the data.

40. The variable long or short affects the rolling time.

- a. true.
- b. false.
- c. cannot tell from the data.

41. The slope of the table affects the rolling time.

- a. true.
- b. false.
- c. cannot tell from the data.

42. The above table suggests that for this experiment the rolling time for hollow cylanders:

- a. is always 10 sec.
- b. does not depend on material.
- c. does not depend on length.
- d. all the above are correct.
- e. none of the above is correct.

43. For this experiment a solid cylander is one which:

- a. is metal.
- b. is 8 cm long.
- c. has a rolling time of 5 seconds.
- d. more than one of the above is correct.
- e. none of the above.

44. For this experiment a metal cylander is one which:

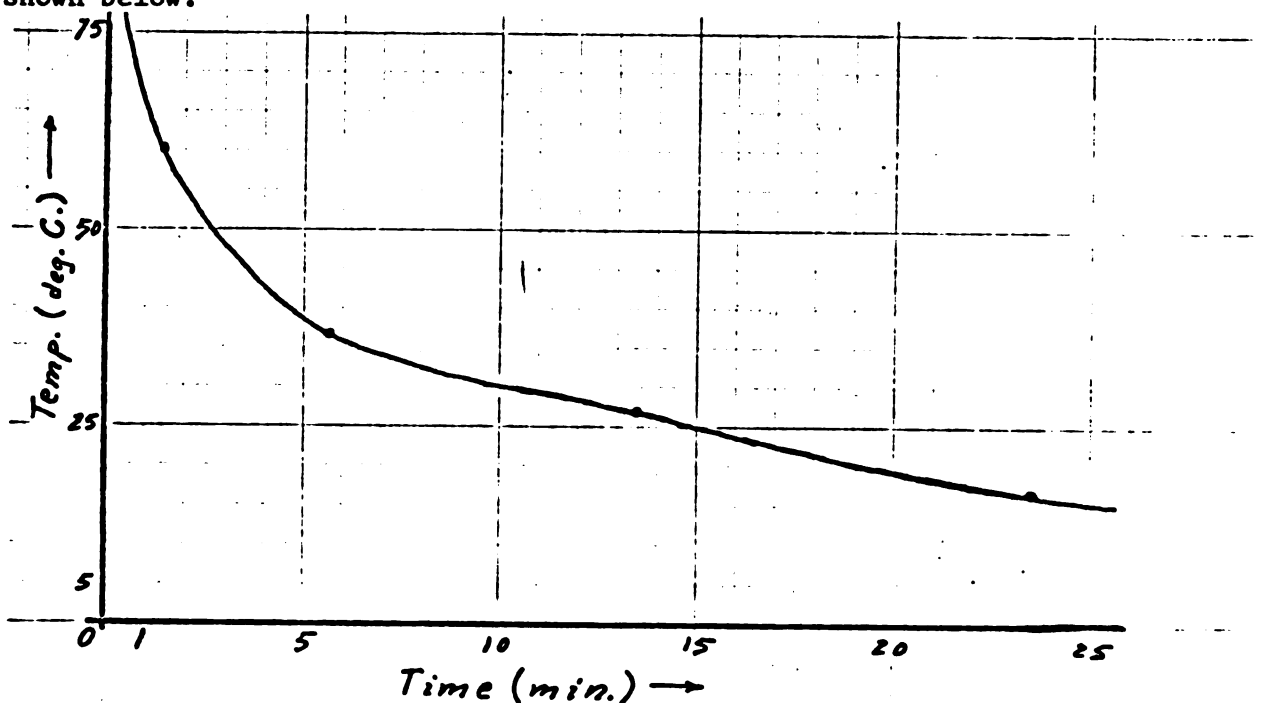
- a. is solid.
- b. is 8 cm long.
- c. has a rolling time of 5 seconds.
- d. more than one of the above is correct.
- e. none of the above is correct.

45. To tell someone how to answer the above questions it would be best to tell them that by "rolling time" I mean the time:

- a. for the cylander to roll the length of the table.
- b. during which gravity is acting on the cylander.
- c. as indicated by my stop watch.
- d. more than one of the above is correct.
- e. none of the above.

Questions 46 to 52 are about the following experiment:

The TV ads claim a certain false teeth cleaner is colored green and the green color disappears when your teeth are clean. To check the time for this reaction the following experiment is performed: The time for the green to disappear from a glass of water is measured at several different temperatures. A graph of the data is shown below.



46. From this information we can say that reaction time
- increases with increases in temperature.
 - decreases with increases in temperature.
 - decreases with decreases in temperature.
 - is not effected by changes in temperature.
 - none of the above.
47. From the graph, how long should it take to clean your false teeth at a temperature of 20 degrees C.
- less than 5 minutes.
 - between 5 and 10 minutes.
 - between 10 and 15 minutes.
 - between 15 and 20 minutes.
 - more than 20 minutes.
48. Where on the graph is the cleaning time most affected by changes in temperature.
- less than 5 minutes.
 - between 5 and 10 minutes.
 - between 10 and 15 minutes.
 - between 15 and 20 minutes.
 - more than 20 minutes.
49. If you needed to clean your false teeth in less than 5 minutes you could use a temperature of:
- zero degrees.
 - 25 degrees.
 - 50 degrees.
 - 75 degrees.
 - more than one of the above.
50. As the temperature increases by 25 degrees, the cleaning time
- increases by between 5 and 10 minutes.
 - decreases by between 5 and 10 minutes.
 - increases by between 10 and 15 minutes.
 - decreases by between 10 and 15 minutes.
 - none of the above.
51. To tell a friend how to measure the cleaning time it would be best to say the cleaning time is the time for:
- the green color to disappear.
 - the chemical reaction to be completed.
 - all the bacteria on the teeth to be killed.
 - all the above are equally good answers.
 - none of the above.

52. Suppose a friend has gone to a lake to swim. He wants to know the water temperature but he has no thermometer. He borrows a tablet of the above brand of denture cleaner, goes out to the end of the dock and drops the tablet in the water. He tries to use the above graph to tell the water temperature. His effort fails, probably because:
- a. the water is too cold to swim in.
 - b. he did not wait long enough.
 - c. he used the wrong amount of water.
 - d. there are no false teeth in the water.
 - e. none of the above.
53. Jean watches a bull fight and decides that bulls charge red objects. To test this idea she should observe a bull in a ring in which:
- a. there is no matador but there are several red objects.
 - b. there is no matador but there are objects of several different colors including some that are red.
 - c. there is a matador who waves a red cape.
 - d. there is a matador who waves capes of different colors including one that is red.
 - e. more than one of the above is correct.
54. When 100 ml of alcohol and 100 ml of water are mixed, somewhat less than 200 ml of solution results. A possible explanation for this observation is:
- a. alcohol evaporates quickly.
 - b. liquids have space between their molecules.
 - c. some liquids cool and contract when mixed.
 - d. more than one of the above is correct.
 - e. none of the above is correct.
55. Suppose a friend dials a number, hands you the phone, and tells you to find out if the store he has called is a hardware store or a grocery store. You are allowed only one question and you cannot use the words "hardware" or "grocery." You should ask:
- a. if they sell can openers.
 - b. what they sell the most of.
 - c. the name of the store.
 - d. the name of the manager.
 - e. none of the above would help.
56. A paper cup is filled with water and held over a lighted candle. Although the flame is very near the cup, the cup does not burn. The reason may be the
- a. cup may have become soaked with water.
 - b. cup may be made of fire proof paper.
 - c. water may be absorbing heat too fast.
 - d. all of the above.
 - d. none of the above.

57. Suppose a space traveler from some distant planet visits you. The people on his planet are just like us except they do not have eyes. He can talk with you, but of course he cannot see. It is your job to tell him what you mean by "sight." It would be best to begin by saying "sight" is:
- a. what I do when I see.
 - b. how I know it is you I am talking to and not someone else.
 - c. how I recognize you and what you are wearing without hearing you speak or touching you.
 - d. the reaction of light on the nerves in the retina of my eye.
 - e. none of the above.

Questions 58 to 60 are about the following experiment:

A scientist wanted to know if a special light bulb is as efficient as sun light. He selected two young bean plants. He placed one plant on his windowsill and the other in a closet. He put his special light bulb in the socket in the closet, turned it on, and closed the door. He returned in three days to see how his plants were doing. He found that the plants had grown exactly the same amount. Therefore he decided his special light bulb is as efficient as sun light.

58. The reason the scientist used 2 plants in the experiment is:
- a. so he could compare the plants.
 - b. in case one plant died, the experiment would not be a failure.
 - c. he really did not need 2 plants.
 - d. his chances of getting a healthy plant were better by using two plants than if he had chosen only one.
 - e. none of the above.
59. By "efficient" the scientist must mean:
- a. the type of chemical reaction that the light source causes.
 - b. the amount of energy delivered to the plant by the light source.
 - c. the ability to cause plant growth.
 - d. all of the above.
 - e. none of the above.
60. This would have been a better experiment if:
- a. more plants had been used.
 - b. the light had been connected to an automatic switch that would turn it on only when the sun was shining.
 - c. the scientist had given the distance from the light bulb to the plant in the closet.
 - d. more than one of the above is correct.
 - e. none of the above.

61. A scientist would say I am doing mechanical work when I pedal my bike but I am not doing mechanical work when I stop pedaling and coast. From this statement alone you might conclude that by "mechanical work" a scientist means that:
- a. motion must occur.
 - b. force must be applied.
 - c. either of the above is enough to mean "mechanical work" to a scientist.
 - d. both force and motion are needed.
 - e. none of the above.
62. A scientist would say I am doing mechanical work when I push a broom but I am not doing work when I stop and lean on the broom while I talk to one of my friends. From this statement alone you might conclude that by "mechanical work" a scientist means that:
- a. motion must occur.
 - b. force must be applied.
 - c. either of the above is enough to mean "mechanical work" to a scientist.
 - d. both force and motion must occur.
 - e. none of the above.
63. If questions 61 and 62 are taken together, you might conclude that by "mechanical work" a scientist means that:
- a. motion must occur.
 - b. force must be applied.
 - c. either of the above is enough to mean "mechanical work" to a scientist.
 - d. both force and motion must occur.
 - e. none of the above.
64. All objects can be bent by some small amount no matter how stiff they are. This idea must be accepted until:
- a. no one believes it any more.
 - b. a scientist says it is no longer true.
 - c. objects are found that bend easily.
 - d. someone finds an object that does not bend.
 - e. none of the above.
65. Mary has a thermometer in her room. Her thermometer is best described as:
- a. an indoor thermometer.
 - b. a mercury-filled glass tube.
 - c. a device for measuring temperature.
 - d. a thermostat
 - e. none of the above.

Questions 66 to 72 are about the following experiment:

A science class decided to check their reaction times thus: each student had to flip a switch as soon as he saw a light flash, heard a buzzer sound, or both. A timer recorded the time it took for each student to react. The data was recorded using the following form:

REACTION TIME DATA		
SEX	STIMULUS	TIME
(B = Boy G = Girl)	(L = light S = sound B = both)	(seconds)

66. Using the information to be recorded in the above table it would be possible to find out:
- who the boy is that has the fastest reaction time.
 - who the girl is that has the fastest reaction time.
 - whether the student with the fastest reaction time is a boy or a girl.
 - more than one of the above.
 - none of the above
67. Using the information to be recorded in the above table it would be possible to find out whether:
- staying up late the night before has any effect on reaction time.
 - the loudness of the buzzer has any effect on reaction time.
 - smoking has any effect on reaction time.
 - more than one of the above.
 - none of the above.
68. Using the information to be recorded in the above table it would be possible to find out whether on the average:
- the light produced quicker reactions than the buzzer.
 - boys required a brighter light to react than girls.
 - time is important.
 - more than one of the above.
 - none of the above.

After the above data had been taken, the class averages were figured.
The results were:

STIMULUS	AVERAGE TIME	
	BOYS .	GIRLS
Light	.17 sec	.15 sec.
Buzzer	.22	.19
Both	.14	.23

69. Who reacted quickest to the buzzer?

- a. boys by .02 sec.
- b. girls by .02 sec.
- c. boys by .03 sec.
- d. girls by .03 sec.
- e. boys by .09 sec.

70. Who reacted quickest to both the light and the buzzer together?

- a. girls by .03 sec.
- b. boys by .05 sec.
- c. girls by .05 sec.
- d. boys by .09 sec.
- e. girls by .09 sec.

71. Did boys react quicker to the light than the girls did to the buzzer?

- a. yes, by .02 sec.
- b. yes, by .09 sec.
- c. no, girls were quicker by .02 sec.
- d. no, girls were quicker by .05 sec.
- e. no, girls were quicker by .08 sec.

72. The term "reaction time" as used above means the time:

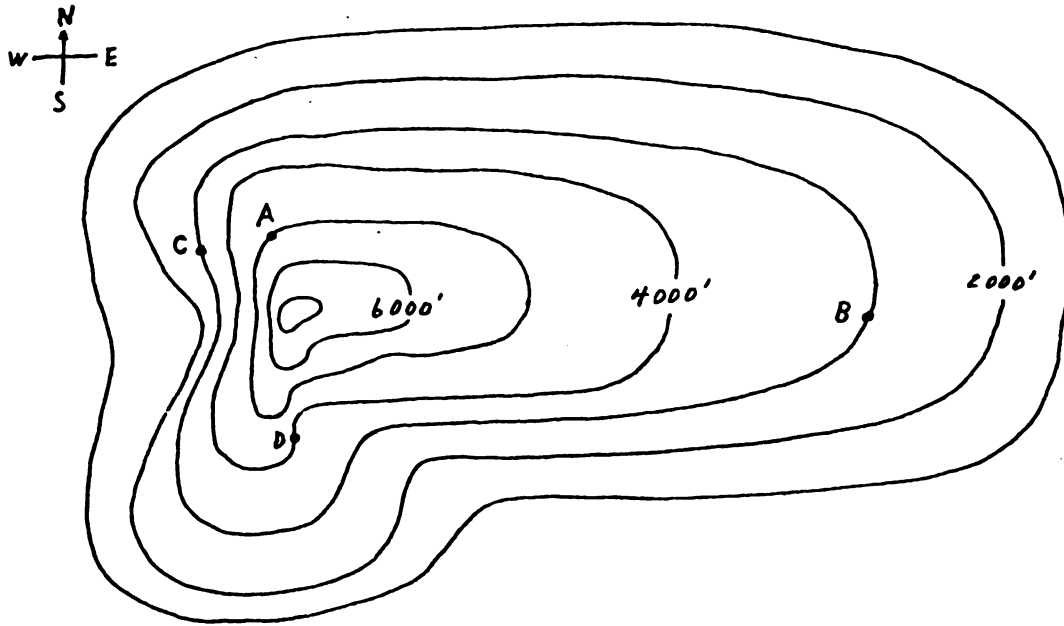
- a. required for nerve impulses to be transmitted.
- b. required for the buzzer to buzz or the light to flash.
- c. required for the buzzer to quit buzzing or the light to quit flashing.
- d. during which the student is deciding how to react
- e. none of the above.

73. You are given a block of wood and a beaker of an unknown liquid. To find out whether the wood will float on the surface of the liquid you should:

- a. find the density of the wood.
- b. find the density of the liquid.
- c. put the block of wood in the liquid and watch it.
- d. put the block of wood in several different kinds of liquids and watch it.
- e. put several different kinds of wood in the unknown liquid and watch them.

74. Which of the following tells most clearly what to do and what to observe:
- a. add 5 ml of sodium hydroxide to 50 ml of grape juice.
 - b. add sodium hydroxide to grape juice and the juice will change color.
 - c. changing the hydroxide concentration of the proper indicator will cause a change in color.
 - d. grape juice contains colored indicators.
 - e. all of the above are quite clear.
75. Suppose it is your job to tell the world what a "mountain" is. Everyone will accept your definition if by using it they can always tell whether or not what they are looking at fits what you mean by a "mountain." It would be best to say, a mountain:
- a. is high.
 - b. is higher than a hill.
 - c. has an altitude of 5000 feet or more.
 - d. requires much work to climb.
 - e. none of the above.
76. A girl removed a lid from a jar by prying on it with the blade of a table knife. From that use of it, you might say a knife is a:
- a. sterling silver object with a sharp edge and a decorated handle.
 - b. stainless steel object about 8 inches long with a thin blade.
 - c. metal object that can be used as a lever to open jars.
 - d. kind of incline plane that reduces the force needed to cut.
 - e. all of the above.

Use the following contour map to answer questions 77 to 80.



77. What is the elevation at point A?

- a. 9000 feet.
- b. 6500 feet.
- c. 6000 feet.
- d. 4000 feet
- e. none of the above.

78. This mountain is steepest on its

- a. north side.
- b. south side.
- c. east side.
- d. west side.
- e. not enough information has been given to tell.

79. Which of the following is at the highest elevation:

- a. A.
- b. B.
- c. C.
- d. D.
- e. not enough information has been given to tell.

80. The "fall line" can be said to be the direction water would flow if it were poured on the ground. From this definition it is safe to say the direction of the fall line at point D would be approximately:

- a. north.
- b. south.
- c. east.
- d. west.
- e. not enough information has been given to tell.

APPENDIX IV - B

FORM A SUBTEST ASSIGNMENTS

	<u>ID*</u> <u>Items</u>	<u>CV*</u> <u>Items</u>	<u>FH*</u> <u>Items</u>	<u>DO*</u> <u>Items</u>
	3	29	1	6
	4	30	2	11
	5	31	7	19
	8	32	9	20
	12	33	10	21
	13	34	15	28
	14	35	16	43
	24	36	17	44
	26	37	18	45
	38	52	22	51
	39	58	23	55
	40	60	25	57
	41	66	27	59
	46	67	42	61
	47	68	53	62
	48		54	63
	49		56	65
	50		64	72
	69			73
	70			74
	71			75
	77			76
	78			80
	<u>79</u>	<u>—</u>	<u>—</u>	<u>—</u>
TOTALS	24	15	18	23

*ID - Interpreting Data

*CV - Controlling Variables

*FH - Formulating Hypothesis

*DO - Defining Operationally

APPENDIX IV - C
ITEM ANALYSIS FORM A

APPENDIX IV - C

ITEM ANALYSIS FORM A

All Values are in Percent

Item	Key	Alt.	SAPA			Diff. Disc.	TRAD.			Diff. Disc.
			U27	M46	L27		U27	M46	L27	
1		A	0	6	25	47	0	0	13	53
		B	13	25	13	37	13	25	25	38
	*	C	75	50	38		63	50	25	
		D	13	19	25		25	25	38	
		E	0	0	0		0	0	0	
<hr/>										
2	*	A	13	25	13	81	25	31	13	72
		B	0	6	13	0	0	25	25	12
		C	13	6	50		13	13	25	
		D	0	13	0		0	0	25	
		E	75	50	25		63	31	13	
<hr/>										
3		A	0	0	0	47	0	13	13	50
		B	0	6	13	50	0	0	25	75
		C	0	6	25		0	0	0	
	*	D	88	44	38		75	63	0	
		E	13	44	25		25	25	50	
<hr/>										
4		A	0	0	25	31	0	25	13	59
		B	0	13	50	87	25	19	0	37
		C	0	6	13		0	25	13	
	*	D	100	81	13		75	25	38	
		E	0	0	0		0	6	25	
<hr/>										
5	*	A	38	19	63	66	25	25	38	63
		B	38	50	0	38	63	44	0	63
		C	0	13	25		0	6	25	
		D	0	0	13		0	6	0	
		E	25	19	0		13	19	25	
<hr/>										
6	*	A	38	25	13	56	0	13	0	47
		B	38	38	63	-25	75	50	38	37
		C	0	6	0		0	6	0	
		D	0	0	13		13	6	13	
		E	25	31	13		13	25	25	

Appendix IV-C cont'd

Item	Key	Alt.	SAPA			Diff. Disc.	TRAD.			Diff. Disc.
			U27	M46	L27		U27	M46	L27	
7		A	25	6	13	69	0	13	0	53
		B	0	38	13	50	0	50	13	50
		C	0	25	50		0	6	38	
	*	D	63	25	13		88	31	38	
		E	16	6	13		13	0	13	
<hr/>										
8		A	38	50	75	91	50	38	50	91
		B	0	6	13	25	0	0	13	25
		C	0	31	13		25	38	0	
		D	38	6	0		0	19	38	
	*	E	25	6	0		25	6	0	
<hr/>										
9		A	13	0	13	50	13	0	13	56
		B	0	19	38	13	0	25	13	38
		C	25	50	38		25	6	25	
	*	D	50	25	0		63	44	25	
		E	13	6	13		0	25	25	
<hr/>										
10		A	13	38	38	78	13	25	13	91
		B	0	6	0	13	13	0	13	25
	*	C	38	13	25		25	6	0	
		D	25	13	13		13	19	50	
		E	25	31	25		38	50	25	
<hr/>										
11	*	A	88	81	75	19	50	44	25	59
		B	0	0	0	13	0	0	13	25
		C	0	0	0		0	6	25	
		D	13	13	25		50	44	38	
		E	0	6	0		0	6	0	
<hr/>										
12		A	13	6	25	69	0	44	25	66
	*	B	38	31	25	13	50	31	25	25
		C	0	6	13		13	0	0	
		D	0	0	0		0	0	25	
		E	38	56	38		38	25	25	
<hr/>										
13		A	0	13	13	59	13	5	13	72
		B	13	19	25	13	13	44	38	0
	*	C	63	25	50		38	19	38	
		D	0	6	13		25	6	13	
		E	25	38	0		13	25	0	

Appendix IV-C cont'd

Item	Key	Alt.	SAPA			Diff. Disc.	TRAD.			Diff. Disc.
			U27	M46	L27		U27	M46	L27	
14		A	13	13	25	56	13	13	50	75
		B	0	13	38	75	0	31	25	50
		C	13	0	25		13	13	13	
		D	0	25	13		25	19	13	
	*	E	75	50	0		50	25	0	
15		A	0	44	50	84	38	50	25	81
		B	75	25	25	0	38	19	50	0
		C	13	13	0		13	6	13	
	*	D	19	19	13		13	25	13	
		E	0	0	13		0	0	0	
16		A	0	0	0	63	0	13	25	56
		B	38	6	38	38	25	25	38	38
		C	0	25	25		0	13	13	
		D	25	13	38		13	6	0	
	*	E	38	56	0		63	44	25	
17		A	0	25	25	59	25	69	38	75
		B	13	0	13	62	0	0	0	38
		C	0	13	38		13	19	25	
	*	D	75	38	13		63	6	25	
		E	13	25	13		0	6	13	
18	*	A	50	38	25	63	25	25	25	75
		B	0	0	0	25	38	38	25	0
		C	13	13	25		0	0	13	
		D	38	31	13		25	25	13	
		E	0	19	38		13	13	25	
19		A	0	6	25	41	0	19	38	59
		B	0	6	25	75	0	13	13	50
	*	C	100	56	25		75	31	25	
		D	0	25	13		25	25	13	
		E	0	6	13		0	13	13	
20		A	25	25	75	69	25	56	38	78
	*	B	38	44	0	38	13	25	25	-12
		C	13	19	13		25	13	25	
		D	0	6	13			0	0	0
		E	25	6	0		38	6	13	

Appendix IV-C cont'd

Item	Key	Alt.	SAPA			Diff.	TRAD.			Diff.
			U27	M46	L27	Disc.	U27	M46	L27	Disc.
21	*	A	0	6	13	28	13	19	0	75
		B	88	88	25	63	50	13	25	25
		C	0	6	13		0	19	13	
		D	0	0	38		38	44	50	
		E	13	0	13		0	6	13	
22		A	0	6	25	91	13	0	38	88
		B	0	13	38	0	25	19	13	-13
		C	100	63	38		63	44	25	
		D	0	0	0		0	19	13	
		E	0	19	0		0	19	13	
23		A	0	0	0	22	0	19	0	34
		B	0	0	13	25	13	0	13	38
		C	0	13	0		0	6	0	
		D	100	69	75		88	63	50	
		E	0	19	13		0	13	38	
24		A	50	19	13	84	13	0	25	81
		B	13	31	38	-12	25	44	0	13
		C	0	6	13		25	13	50	
		D	25	31	13		25	13	13	
		E	13	13	25		13	31	0	
25		A	0	0	13	81	0	13	13	69
		B	63	44	38	0	25	25	25	37
		C	13	13	13		13	19	13	
		D	25	13	25		50	31	13	
		E	0	31	13		13	6	25	
26	*	A	38	19	50	69	25	19	13	81
		B	13	19	25	-12	38	6	13	12
		C	38	31	25		38	44	25	
		D	0	13	0		0	13	0	
		E	13	19	0		0	13	38	
27	*	A	0	19	0	44	13	25	13	88
		B	63	56	50	13	25	6	13	12
		C	13	13	13		25	25	13	
		D	25	0	25		13	38	25	
		E	0	13	13		25	6	25	

Appendix IV-C cont'd

Item	Key	Alt.	SAPA			Diff. Disc.	TRAD.			Diff. Disc.
			U27	M46	L27		U27	M46	L27	
28	*	A	88	25	25	59	38	38	0	72
		B	13	25	25	63	38	6	38	38
		C	0	0	13		0	31	0	
		D	0	19	0		25	13	25	
		E	0	31	38		0	13	25	
29		A	25	19	0	78	13	13	13	81
		B	25	19	25	37	13	19	50	38
		C	0	19	50		13	19	0	
	*	D	50	13	13		38	19	0	
		E	0	31	13		25	31	25	
30	*	A	13	6	0	78	0	0	0	63
		B	0	25	38	-38	50	44	13	37
		C	50	19	13		13	13	25	
		D	38	38	50		25	31	38	
		E	0	13	0		13	13	13	
31		A	38	25	25	81	38	19	38	94
		B	13	19	75	25	25	38	0	-13
		C	0	13	0		0	6	38	
		D	25	19	0		38	31	0	
	*	E	25	25	0		0	6	13	
32		A	0	21	11	35	30	28	10	58
		B	0	0	11	22	0	0	10	30
	*	C	89	53	67		60	39	30	
		D	0	16	11		10	22	40	
		E	11	11	0		0	11	10	
33	*	A	11	11	11	30	0	11	20	37
		B	67	79	56	11	90	50	60	30
		C	0	0	33		10	6	0	
		D	0	5	0		0	11	20	
		E	22	5	0		0	22	0	
34		A	22	16	33	92	10	17	30	97
		B	56	26	44	-11	40	67	10	10
		C	22	42	11		30	11	30	
		D	0	5	0		10	6	30	
	*	E	0	11	11		10	0	0	

Appendix IV-C cont'd

Item	Key	Alt.	SAPA		L27	Diff. Disc.	TRAD.			Diff. Disc.
			U27	M46			U27	M46	L27	
35	*	A	11	5	22	41	0	17	40	66
		B	44	74	44	0	50	39	10	40
		C	11	0	11		0	6	20	
		D	11	11	0		10	22	20	
		E	22	11	22		40	17	10	
36	*	A	33	16	22	68	20	28	30	66
		B	0	32	22	34	0	6	0	30
		C	56	26	22		60	22	30	
		D	0	16	22		20	17	20	
		E	11	11	11		0	22	10	
37	*	A	11	26	33	70	0	39	20	71
		B	33	16	11	22	20	22	40	50
		C	44	26	22		60	22	10	
		D	11	16	22		20	6	20	
		E	0	16	11		0	6	10	
38	*	A	33	26	56	46	10	28	30	39
		B	56	63	33	23	90	61	30	60
		C	11	11	11		0	11	30	
		D	0	0	0		0	0	0	
		E	0	0	0		0	0	10	
39	*	A	100	84	67	16	80	72	30	37
		B	0	11	22	33	20	22	50	50
		C	0	5	0		0	6	10	
		D	0	0	0		0	0	0	
		E	0	0	0		0	0	0	
40	*	A	0	21	33	41	10	28	70	55
		B	100	53	33	67	80	50	0	80
		C	0	26	22		10	22	30	
		D	0	0	11		0	0	0	
		E	0	0	0		0	0	0	
41	*	A	11	11	56	32	10	28	50	50
		B	0	5	22	78	0	22	20	70
		C	89	84	11		90	44	20	
		D	0	0	0		0	0	0	
		E	0	0	11		0	0	0	

Appendix IV-C cont'd

Item	Key	Alt.	SAPA			Diff. Disc.	TRAD.			Diff. Disc.
			U27	M46	L27		U27	M46	L27	
42		A	44	58	56	70	50	33	30	79
		B	0	5	11	22	0	0	0	40
		C	0	5	0		0	28	20	
	*	D	44	26	22		50	11	10	
		E	11	5	11		0	28	30	
<hr/>										
43		A	0	5	22	38	0	11	30	63
		B	0	0	22	67	10	11	0	30
	*	C	100	58	33		50	39	20	
		D	0	16	22		40	28	20	
		E	0	21	0		0	11	20	
<hr/>										
44		A	0	0	11	41	0	17	10	68
		B	0	0	11	78	0	6	20	60
		C	0	11	33		20	17	30	
		D	11	21	33		20	28	30	
	*	E	89	68	11		60	33	0	
<hr/>										
45	*	A	67	79	33	35	60	67	30	45
		B	0	5	22	34	0	11	10	30
		C	22	5	22		0	11	30	
		D	0	11	11		40	6	30	
		E	11	0	11		0	6	0	
<hr/>										
46		A	44	26	56	70	30	22	30	63
	*	B	33	32	22	11	50	44	10	40
		C	22	37	0		20	11	0	
		D	0	0	11		0	6	40	
		E	0	5	0		0	17	20	
<hr/>										
47		A	0	5	0	54	0	17	10	58
		B	11	16	0	56	0	6	0	40
		C	0	11	44		20	28	30	
	*	D	67	53	11		70	33	30	
		E	22	16	33		10	17	20	
<hr/>										
48	*	A	78	32	11	62	80	22	30	61
		B	0	16	22	67	0	33	10	50
		C	0	16	0		10	11	0	
		D	0	16	22		0	0	20	
		E	22	21	33		10	33	30	

Appendix IV-C cont'd

Item	Key	Alt.	SAPA			Diff. Disc.	TRAD.			Diff. Disc.
			U27	M46	L27		U27	M46	L27	
49		A	0	11	0	89	0	6	20	87
		B	0	37	11	-11	0	17	20	10
		C	11	16	22		30	28	0	
		D	78	32	33		50	39	40	
	*	E	11	5	22		20	11	10	
<hr/>										
50		A	11	26	44	92	20	17	10	92
		B	33	37	11	11	20	39	10	0
		C	44	21	22		30	33	40	
		D	0	5	11		20	6	20	
	*	E	11	11	0		10	6	10	
<hr/>										
51	*	A	78	47	33	49	10	17	10	87
		B	11	5	11	45	0	11	20	0
		C	0	11	0		10	17	10	
		D	11	37	33		80	50	20	
		E	0	0	0		00	6	10	
<hr/>										
52		A	0	11	0	70	10	11	10	87
		B	22	5	22	44	10	22	20	20
	*	C	44	37	0		30	6	10	
		D	11	0	22		40	28	10	
		E	22	47	44		10	33	40	
<hr/>										
53		A	11	11	22	76	30	17	0	74
	*	B	33	32	0	33	20	22	40	-20
		C	0	16	22		10	17	20	
		D	22	11	22		0	6	10	
		E	33	32	22		40	39	20	
<hr/>										
54		A	11	16	11	70	10	22	20	76
		B	11	21	22	11	0	6	30	10
		C	33	26	11		20	39	10	
	*	D	44	21	33		30	22	20	
		E	0	16	11		40	11	10	
<hr/>										
55		A	0	11	11	68	0	0	20	53
	*	B	56	21	33	23	44	90	44	80
		C	33	47	44		10	39	40	
		D	0	0	0		0	11	0	
		E	11	21	0		0	6	20	

Appendix IV-C cont'd

Item	Key	Alt.	SAPA			Diff. Disc.	TRAD.			Diff. Disc.
			U27	M46	L27		U27	M46	L27	
56		A	22	26	33	78	30	11	10	63
		B	22	0	22	22	0	0	20	10
		C	22	37	22		30	50	30	
	*	D	33	21	11		40	39	30	
		E	0	16	0		0	0	0	
<hr/>										
57		A	0	0	22	62	0	6	30	55
		B	11	5	0	56	0	17	10	50
	*	C	67	37	11		60	56	10	
		D	22	42	56		30	22	30	
		E	0	16	0		10	0	10	
<hr/>										
58	*	A	100	84	56	19	100	67	70	24
		B	0	0	0	44	0	6	0	30
		C	0	0	11		0	6	10	
		D	0	0	22		0	11	10	
		E	0	16	0		0	11	0	
<hr/>										
59		A	0	11	33	89	0	6	30	76
		B	11	21	22	11	20	39	30	10
	*	C	22	5	11		30	22	20	
		D	44	11	0		50	6	20	
		E	11	21	11		0	28	0	
<hr/>										
60		A	22	0	22	84	10	17	0	61
		B	22	68	33	44	30	17	10	10
		C	0	0	22		0	11	30	
	*	D	44	11	0		40	44	30	
		E	11	21	11		20	11	20	
<hr/>										
61		A	11	11	11	76	10	17	20	79
	*	B	33	21	22	11	20	22	20	0
		C	11	5	11		10	17	20	
		D	44	63	33		60	44	40	
		E	0	0	11		0	0	0	
<hr/>										
62	*	A	44	26	22	70	40	11	0	84
		B	22	21	33	22	10	17	30	40
		C	11	11	33		10	11	30	
		D	22	37	0		40	50	20	
		E	0	5	0		0	11	20	

Appendix IV-C cont'd

Item	Key	SAPA				Diff. Disc.	TRAD.			Diff. Disc.
		Alt.	U27	M46	L27		U27	M46	L27	
63		A	11	11	11	49	10	0	10	47
		B	11	11	11	34	0	11	10	50
		C	11	26	33		10	39	30	
	*	D	67	53	33		80	50	30	
		E	0	0	0		0	0	20	
64		A	0	0	0	22	0	6	10	53
		B	0	0	11	56	0	28	20	90
		C	0	16	11		0	17	20	
	*	D	100	84	44		100	39	10	
		E	0	0	22		0	11	30	
65		A	11	21	67	51	10	44	30	55
		B	0	0	0	67	0	0	20	60
	*	C	78	53	11		90	28	30	
		D	11	16	22		0	22	20	
		E	0	0	0		0	6	0	
66		A	0	0	11	32	20	6	10	45
		B	0	0	11	33	0	0	0	20
	*	C	89	63	56		60	61	40	
		D	11	16	22		20	17	20	
		E	0	21	0		0	11	20	
67		A	0	11	33	54	0	11	20	76
		B	22	21	33	67	30	39	30	40
		C	0	0	11		0	0	0	
		D	11	11	22		20	33	30	
	*	E	67	58	0		50	17	10	
68	*	A	89	32	56	49	40	17	10	79
		B	0	11	0	33	0	17	30	30
		C	0	16	44		30	33	30	
		D	0	32	0		10	17	20	
		E	11	11	0		20	11	0	
69		A	0	0	11	41	0	6	10	45
		B	11	16	0	45	10	17	30	60
		C	0	21	33		0	6	0	
	*	D	89	53	44		90	50	30	
		E	0	5	11		0	17	20	

Appendix IV-C cont'd

Item	Key	Alt.	SAPA			Diff. Disc.	TRAD.			Diff. Disc.
			U27	M46	L27		U27	M46	L27	
70		A	0	11	0	32	0	6	50	50
		B	0	0	0	33	0	6	0	70
		C	0	5	11		10	22	20	
	*	D	89	63	56		90	44	20	
		E	11	16	22			0	17	0
71	*	A	56	42	22	59	80	44	20	53
		B	11	11	44	34	10	17	20	60
		C	33	42	33		10	22	30	
		D	0	0	0		0	6	10	
		E	0	0	0		0	0	10	
72		A	89	53	33	92	60	56	10	97
		B	0	5	22	0	0	17	20	10
		C	0	5	11		0	11	10	
		D	11	16	33		30	6	50	
	*	E	0	16	0		10	0	0	
73		A	0	11	22	92	0	11	50	55
		B	0	0	0	78	10	6	10	70
	*	C	100	63	22		90	33	20	
		D	0	0	33		0	28	0	
		E	0	21	22		0	11	10	
74		A	11	21	22	73	0	17	20	79
	*	B	56	21	11	45	50	6	20	30
		C	22	21	22		30	22	20	
		D	0	5	11		0	0	0	
		E	11	26	22		20	39	30	
75		A	0	0	22	73	0	11	0	79
		B	22	32	11	0	30	22	30	0
	*	C	33	21	33		10	33	10	
		D	0	0	11		40	11	30	
		E	44	42	22		10	6	20	
76		A	0	0	0	46	0	6	0	39
		B	0	5	22	45	0	6	10	30
	*	C	89	42	44		80	56	50	
		D	0	26	33		10	6	20	
		E	11	21	0		0	11	10	

Appendix IV-C cont'd

Item	Key	Alt.	SAPA			Diff.	TRAD.			Diff.
			U27	M46	L27	Disc.	U27	M46	L27	Disc.
77	*	A	0	0	0	65	0	11	10	58
		B	0	21	22	67	10	6	20	50
		C	22	37	33		20	17	50	
		D	0	11	33		0	6	0	
	*	E	78	26	11		60	50	10	
<hr/>										
78		A	0	5	0	51	0	6	0	61
		B	0	0	22	23	0	0	30	90
		C	11	11	22		0	44	10	
		D	67	42	44		90	33	0	
		E	22	37	11		0	6	50	
<hr/>										
79	*	A	78	63	33	41	90	56	50	37
		B	0	0	11	45	0	11	10	40
		C	0	5	22		0	0	10	
		D	0	11	0		0	6	20	
		E	22	16	33		0	17	0	
<hr/>										
80	*	A	0	0	0	81	0	22	10	87
		B	44	26	67	-11	70	28	30	
		C	22	11	33		0	28	0	
		D	11	16	0		0	0	20	
		E	22	42	0		20	11	30	

APPENDIX IV-D

TSPT FORM C

THE SCIENCE PROCESSES TEST

FORM B C

PART I

MICHIGAN STATE UNIVERSITY

Use the answer sheet provided.

Please do not make any marks on this test booklet.

1. A stick 100 centimeters long is slowly pushed over the edge of a table as shown. About how far over the edge do you think the end will reach before the stick will fall?

- A. 25 cm.
- B. 50 cm.
- C. 75 cm.
- D. 100 cm.



Figure 1.

2. The picture that would show the answer to the above question best would be a picture showing
- A. The stick balanced on my finger.
 - B. How thick the stick is.
 - C. The stick after it has fallen off the table.
 - D. Any of the above pictures would give the answer.

Questions 3, 4 and 5 use the following set up:

Several identical sticks are stacked on top of each other and extended over the edge of a table in such a way as to give the greatest possible overhang (see below).



Figure 2

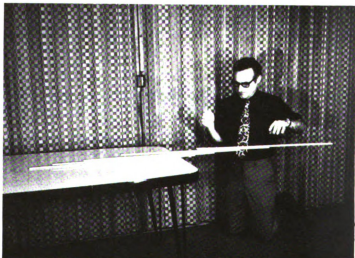


Figure 3

The relation between greatest overhang and number of sticks is graphed below:

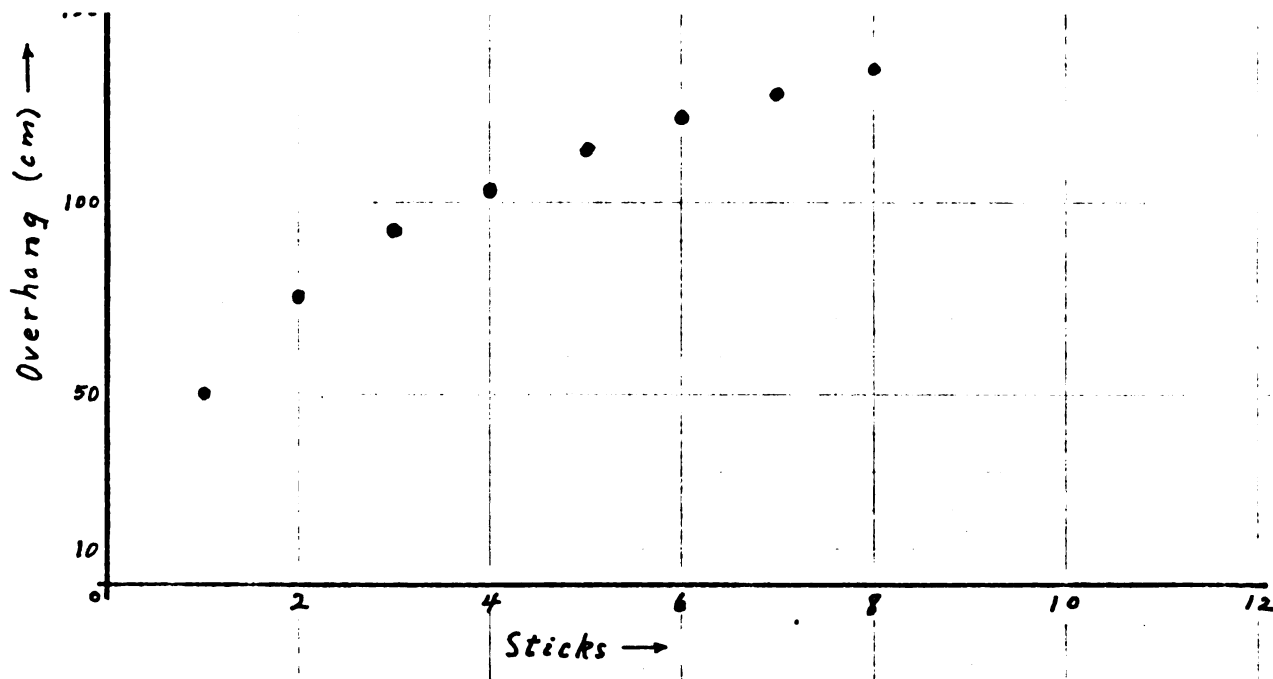


Figure 4

3. The greatest overhang you could get using 5 sticks would be about:
 - A. 1 cm.
 - B. 49.9 cm.
 - C. 99 cm.
 - D. 112 cm.
4. The smallest number of sticks you would need to get an overhang of 100 cm is:
 - A. 1 stick
 - B. 2 sticks
 - C. 3 sticks
 - D. 4 sticks
5. Using 10 sticks you could get a maximum overhang of:
 - A. More than 150 cm.
 - B. Between 130 and 150 cm.
 - C. Between 110 and 130 cm.
 - D. Less than 110 cm.
6. If you needed to tell someone what I mean above by "overhang" so that they could measure the overhang, you could say "overhang" means the distance from the:
 - A. Far end of the top stick to the center of gravity of the system.
 - B. Far end of the top stick to the edge of the table.
 - C. Far end of the top stick to the far end of the next stick.
 - D. Center of gravity of the top stick to the edge of the table.

Questions 7 through 10 are about frames A and B shown in Figure 5.

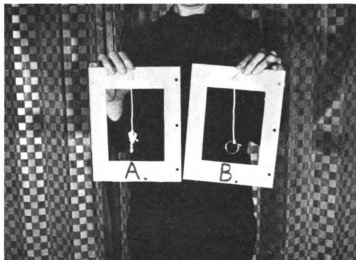


Figure 5

In Figure 6 Frame B has been turned upside down.

7. From Figures 5 and 6 only, it is possible that:
- A. The string in Frame A will remain straight when Frame A is turned upside down.
 - B. The string in Frame A will bend when Frame A is turned upside down.
 - C. The string in Frame B is held straight by a fine thread fastened to the bottom of the frame.
 - D. All of the above may be true.

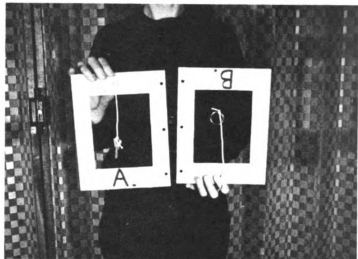


Figure 6

8. Using Figures 5, 6, and 7, the best evidence you have that Frame A is different from Frame B is that when the frames were turned upside down:
- A. The key in Frame A fell.
 - B. The ring in Frame B did not fall.
 - C. Either A. or B. above is evidence there is a difference.
 - D. Both A. and B. above are needed to have evidence for a difference.

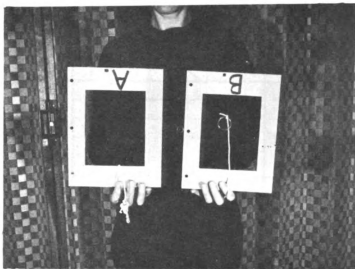


Figure 7

9. Using Figures 5, 6, 7 and 8, the best conclusion is:
- A. The string in Frame A is not as stiff as the string in Frame B.
 - B. The string in Frame A is made of a stiff wire that is now bent.
 - C. The string in Frame B is made of a stiff wire that is now bent.
 - D. The string in Frame B is held up by a strong magnet hidden behind the frame.

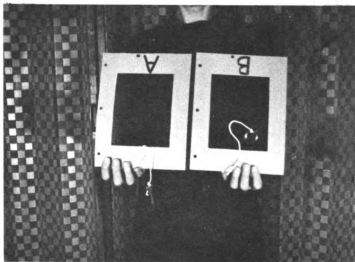


Figure 8

10. My name for the special string used in Frame B above is "Wyrstring." Suppose a friend phoned you to find out if a piece of string he found was Wyrstring. To tell him it would be most helpful to know:
- A. How big around his string is.
 - B. How long his string is.
 - C. How stiff his string is.
 - D. What his string is made of.

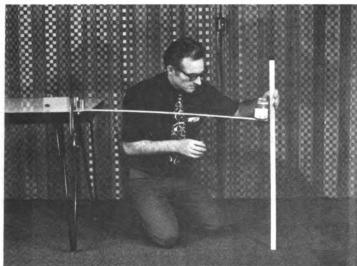


Figure 9

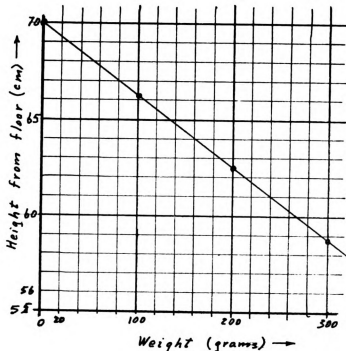


Figure 10

The graph on the right was made by setting weights on the end of the stick which is clamped to the table as shown in Figure 9.

11. The can of soup shown above bends the stick to 66 cm from the floor. The can weighs about:
 - A. 67.5 grams
 - B. 110 grams
 - C. 255 grams.
 - D. 315 grams.
12. A weight of 65 grams should bend the stick to about
 - A. 57.5 cm from the floor.
 - B. 64 cm from the floor.
 - C. 67.5 cm from the floor.
 - D. 135 cm from the floor.
13. The weight that would bend the stick to 72 cm from the floor would weigh about:
 - A. 55 gm.
 - B. 67.5 gm.
 - C. 215 gm.
 - D. None of the above is correct.

Questions 14 through 17 are about the following experiment.

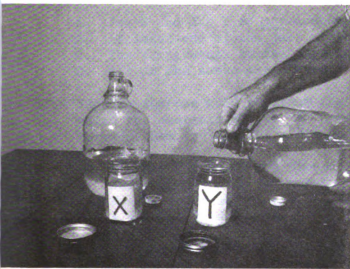


Figure 11

The two jars shown are filled full to the brim.
The lids are screwed on tight.

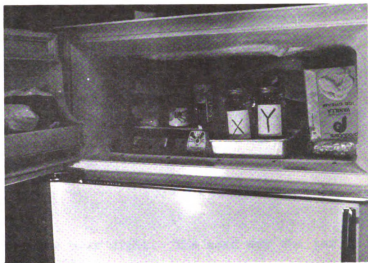


Figure 12

The jars are put in a pan and placed in a freezer.
The temperature inside the freezer is 0 degrees F.

14. Two hours later it is found that neither of the liquids in the jars is frozen. John says the reason is that two hours is not long enough to freeze water. Tom says the reason is that the liquid in the jars is not water, but is some kind of antifreeze.
- A. Probably John is right and Tom is wrong.
 - B. Probably John is wrong and Tom is right.
 - C. Both John and Tom could be right.
 - D. It is unreasonable that either John or Tom is right.
15. One might expect both jars to be frozen solid one day later because:
- A. Water freezes at temperatures below 32 degrees F.
 - B. The liquids in the jars look like water.
 - C. One day in a freezer should be long enough to freeze water.
 - D. All of the above are true.

Next day when the freezer is opened, jar "Y" is found broken and its contents are frozen solid: The contents of jar "X" is still liquid.

16. You now know that:
- A. Jar X and Jar Y do not contain the same kind of liquids.
 - B. At least one of the jars contains water.
 - C. The temperature of the jars is different.
 - D. All of the above are correct.
17. Suppose someone told you that the contents of jar "Y" behaved like a "Cronon" while the contents of jar "X" behaved like a "non Cronon." To show the difference between them you can say a "Cronon" is:
- A. A chemical, but a non Cronon is not.
 - B. Just another name for water.
 - C. Easier to freeze than a non Cronon.
 - D. A special kind of jar.
18. A jar of water was left in the above freezer over night. The next day the jar is broken and the water is frozen. You know that:
- A. Water freezes easier than a Cronon.
 - B. Water behaves like a Cronon.
 - C. Cronons are made of water.
 - D. The water was cold before it was put in the freezer.
19. A bottle of alcohol was left in the above freezer over night. It was not frozen. You now know that:
- A. Alcohol is a Cronon.
 - B. Alcohol is a non Cronon.
 - c. Non Cronons are made of alcohol
 - D. The alcohol was warm before it was put in the freezer.
20. Dick claims that any glass bottle will break when the water it contains freezes. To test his idea, he puts three bottles in the freezer. A is empty, B is half full of water and C is brim full of water. If Dick is correct, which bottles will break?
- A. A only.
 - B. A and B.
 - C. B and C.
 - D. C only.

Questions 21 to 33 are about the cylinders shown below:

The Cylinders shown on the right are described as follows:

Metal cylinders: A, B, C, D.

Plastic cylinders: E, F, G, H.

Short cylinders: A, B, E, F

Long cylinders: C, D, G, H.

Solid cylinders: A, C, E, G.

Hollow cylinders: B, D, F, H.

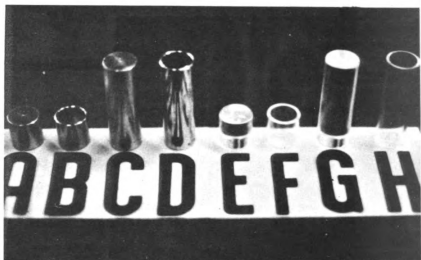


Figure 13

Some or all of the above variables will probably affect the time it takes for the cylinders to roll the length of the table shown. The slope of the table is kept constant by the books under the table legs.

21. By comparing rolling times for cylinders A and E, you could test the effect of the variable:

- A. Solid or hollow.
- B. Long or short.
- C. Metal or plastic.
- D. Amount of slope.

22. By comparing rolling times for cylinders F and H you could test the effect of the variable:

- A. Solid or hollow.
- B. Long or short.
- C. Metal or plastic.
- D. Amount of slope.



Figure 14

23. It was found that the rolling time was exactly the same for cylinders B and D. From this information alone, the variable that does not affect the rolling time is:
- Solid or hollow.
 - Long or short.
 - Metal or plastic.
 - Amount of slope.
24. Which cylinders could you use to tell if a hollow cylinder rolls at a different rate than a solid cylinder?
- A and D.
 - A and F.
 - C and D.
 - C and H.
25. Which cylinders could you use to tell if a metal cylinder rolls at a different rate than a plastic cylinder?
- A and F.
 - A and H.
 - C and G.
 - C and H.

Rolling time for the above cylinders is given in the following table. Use it to answer questions 26 to 33.

<u>Cylinder</u>	<u>Material</u>	<u>Length</u>	<u>Type</u>	<u>Time</u>
A	metal	2 cm	solid	5 sec.
B	metal	2	hollow	10
C	metal	8	solid	5
D	metal	8	hollow	10
E	plastic	2	solid	5
F	plastic	2	hollow	10
G	plastic	8	solid	5
H	plastic	8	hollow	10

26. The material from which the cylinder is made affects the rolling time.
- True.
 - False.
 - Cannot tell from the data.
27. The variable solid or hollow affects the rolling time.
- True.
 - False.
 - Cannot tell from the data.
28. The variable long or short affects the rolling time.
- True.
 - False.
 - Cannot tell from the data.

29. The slope of the table affects the rolling time.
- A. True.
 - B. False.
 - C. Cannot tell from the data.
30. The above table shows that for this experiment the rolling time for hollow cylinders:
- A. Is always the same
 - B. Depends on how long the cylinders are
 - C. Is different for metal than for plastic
 - D. Depends on the slope of the table
31. For this experiment a solid cylinder is one which:
- A. Is metal.
 - B. Has a rolling time of 5 seconds.
 - C. More than one of the above is correct.
 - D. None of the above is correct.
32. For this experiment a metal cylinder is one which:
- A. Is solid.
 - B. Has a rolling time of 5 seconds.
 - C. More than one of the above is correct.
 - D. None of the above is correct.
33. To tell someone how to answer the above questions it would be best to tell them that by "rolling time" I mean the time:
- A. For the cylinder to roll the length of the table.
 - B. During which gravity is pulling on the cylinder.
 - C. As shown by my stop watch.
 - D. Needed for me to start and stop my stopwatch.

End Part I

THE SCIENCE PROCESSES TEST

FORM ~~A~~ C

PART II

MICHIGAN STATE UNIVERSITY

Use the answer sheet provided.

Please do not make any marks on this test booklet.

Questions 34 to 39 are about the following experiment:

The TV ads say that their falseteeth cleaner colors the water green. When the color goes away, your teeth are clean. To find the time needed for the color to go away I got several glasses of water. Each at a different temperature. I put a tablet of falseteeth cleaner in each glass and measured the time needed for the green color to go away. A graph of the data is shown below.

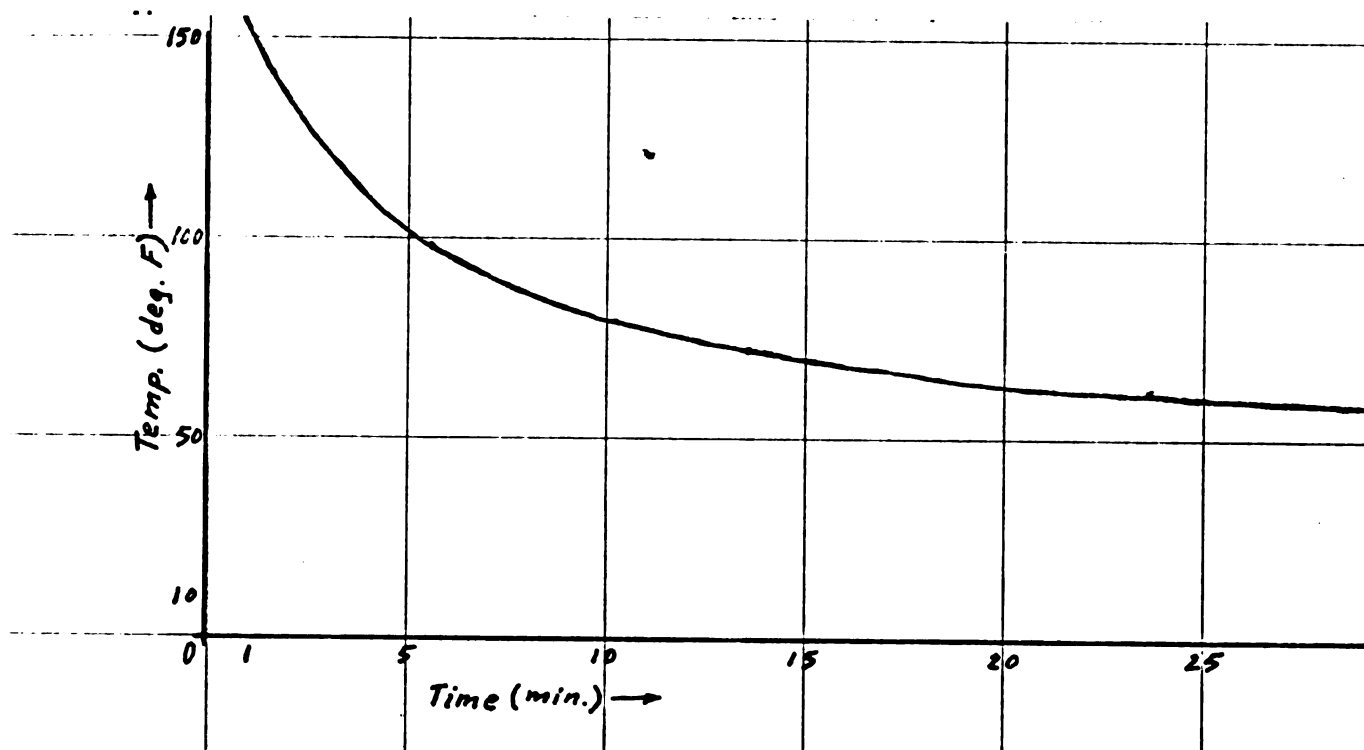


Figure 15

34. From this graph we can say that for a higher temperature, the time needed for the green color to go away is:
- A. Greater.
 - B. Less.
 - C. Not affected by changes in temperature.
 - D. Not enough information is given to tell.
35. From the graph, how long should it take to clean your falseteeth at a temperature of 70 degrees F.
- A. 5 minutes or less.
 - B. Close to 10 minutes.
 - C. Close to 15 minutes.
 - D. 20 minutes or more.
36. Where on the graph is the cleaning time most affected by changes in temperature.
- A. High temperatures.
 - B. Low temperatures.
 - C. Cleaning time is always affected the same amount by a change of temperature.
 - D. Cleaning time will always be the same.

37. If you needed to clean your falseteeth in 5 minutes, you should use a temperature of:
- A. Less than 60 degrees.
 - B. Between 60 and 75 degrees.
 - C. Between 75 and 100 degrees.
 - D. Greater than 100 degrees.
38. To tell a friend how to measure the cleaning time it would be best to say the cleaning time is the time for:
- A. The green color to go away.
 - B. The chemical reaction to be completed.
 - C. All the bacteria on the teeth to be killed.
 - D. All the above are equally good answers.
39. Suppose a friend has gone to a lake to swim. He wants to know the water temperature but he has no thermometer. He borrows a tablet of the above falseteeth cleaner, and drops it into the lake. He tries to use the above graph to tell the water temperature. His effort fails, probably because:
- A. The water is too cold to swim in.
 - B. He did not wait long enough.
 - C. He used the wrong amount of water.
 - D. There are no falseteeth in the water.
40. Jean watches a bull fight and decides that bulls charge red objects. To test this idea she should observe a bull in a ring in which:
- A. There is no matador but there are several red objects.
 - B. There is no matador but there are objects of several different colors including some that are red.
 - C. There is a matador who waves a red cape.
 - D. There is a matador who waves capes of different colors including one that is red.
41. Suppose a friend dials a number, hands you the phone, and tells you to find out if the store he has called is a hardware store or a grocery store. You can ask only one of the following questions. You should ask:
- A. If they sell can openers.
 - B. If they sell light bulbs.
 - C. What they sell the most of.
 - D. How old their store is..

42. Suppose a space traveler from some distant planet visits you. The people on his planet are just like us except they do not have eyes. He can talk with you, but of course he cannot see. It is your job to tell him what you mean by "sight." It would be best to begin by saying "sight" is:
- A. What I do when I see.
 - B. How I know it is you I am talking to and not someone else.
 - C. How I know you and what you are wearing without hearing you speak or touching you.
 - D. The reaction of light on the nerves in the retina of my eye.

Questions 43 to 45 are about the following experiment:

A scientist wanted to know if a special light bulb is as "efficient" as sunlight. He selected two young bean plants. He placed one plant on his windowsill and the other in a closet. He put his special light bulb in the socket in the closet, turned it on, and closed the door. He returned in three days to see how his plants were doing. He found that the plants had grown exactly the same amount. Therefore he decided his special light bulb is as "efficient" as sunlight.

43. The reason the scientist used 2 plants in the experiment is:
- A. So he could compare the plants.
 - B. In case one plant died, the experiment would not be a failure.
 - C. He really did not need 2 plants.
 - D. His chances of getting a healthy plant were better by using two plants than if he had chosen only one.
44. By "efficient" the scientist must mean:
- A. The type of chemical reaction that the light source causes.
 - B. The amount of energy delivered to the plant by the light source.
 - C. The ability to cause plant growth.
 - D. All of the above.
45. This would have been a better experiment if:
- A. More plants had been used.
 - B. The light had been connected to an automatic switch that would turn it on only when the sun was shining.
 - C. The scientist had given the distance from the light bulb to the plant in the closet.
 - D. More than one of the above would help.

46. A scientist would say I am doing "work" when I pedal my bike but I am not doing "work" when I stop pedaling and coast. From this statement alone you might conclude that by "work" a scientist means that:
- A. Motion must occur.
 - B. Force must be applied.
 - C. Either of the above is enough to mean "work" to a scientist.
 - D. Both force and motion are needed.
47. A scientist would say I am doing "work" when I push a broom but I am not doing work when I stop and lean on the broom while I talk to one of my friends. From this statement alone you might conclude that by "work" a scientist means that:
- A. Motion must occur.
 - B. Force must be applied.
 - C. Either of the above is enough to mean "work" to a scientist.
 - D. Both force and motion must occur.
48. If questions 46 and 47 are taken together, you might conclude that by "work" a scientist means that:
- A. Motion must occur.
 - B. Force must be applied.
 - C. Either of the above is enough to mean "work" to a scientist.
 - D. Both force and motion must occur.
49. All objects can be bent by some small amount no matter how stiff they are. This idea must be accepted until:
- A. No one believes it any more.
 - B. A scientist says it is no longer true.
 - C. Objects are found that bend easily.
 - D. Someone finds an object that does not bend.
50. Mary has a thermometer in her room. Her thermometer is best described as:
- A. An indoor thermometer.
 - B. A glass tube containing a colored liquid.
 - C. A device for measuring temperature.
 - D. A thermostat.

Questions 51 to 55 are about the following experiment:

A science class decided to check their reaction times. Each student was asked to flip a switch as soon as he saw a light flash, heard a buzzer sound, or both. A timer recorded the time it took for each student to react. The data was recorded using this data table:

REACTION TIME DATA		
SEX	STIMULUS	TIME
(B = Boy G = Girl)	(L = light S = sound B = both)	(seconds)

51. Using the information to be recorded in the above table it would be possible to find out:
- A. Who the boy is that has the fastest reaction time.
 - B. Who the girl is that has the fastest reaction time.
 - C. Whether the student with the fastest reaction time is a boy or a girl.
 - D. More than one of the above.
52. Using the information to be recorded in the data table it would be possible to find out whether on the average:
- A. The light produced quicker reactions than the buzzer.
 - B. Boys required a brighter light to react than girls.
 - C. Time is important.
 - D. More than one of the above.

After the above data had been taken, the class averages were figured. The results were:

STIMULUS	AVERAGE TIME	
	Boys	Girls
Light	.17 sec	.15 sec.
Buzzer	.22	.19
Both	.14	.23

53. Who reacted more quickly to the buzzer?
- Boys by .02 sec.
 - Girls by .02 sec.
 - Boys by .03 sec.
 - Girls by .03 sec.
 - Boys by .09 sec.
54. Who reacted more quickly to both the light and the buzzer together?
- Girls by .03 sec.
 - Boys by .05 sec.
 - Girls by .05 sec.
 - Boys by .09 sec.
 - Girls by .09 sec.
55. Did boys react more quickly to the light than the girls did to the buzzer?
- Yes, by .02 sec.
 - Yes, by .09 sec.
 - No, girls were quicker by .02 sec.
 - No, girls were quicker by .05 sec.
 - No, girls were quicker by .08 sec.
56. You are given a block of wood and a glass full of an unknown liquid. To find out whether the wood will float on the surface of the liquid you should:
- Find the density of the wood.
 - Find the density of the liquid.
 - Put the block of wood in the liquid and watch it.
 - Put the block of wood in several different kinds of liquids and watch it.
 - Put several different kinds of wood in the unknown liquid and watch them.
57. A girl removed a lid from a jar by prying on it with the blade of a table knife. From that use of it, you might say a knife is a:
- Sterling silver object with a sharp edge and a decorated handle.
 - Stainless steel object about 8 inches long with a thin blade.
 - Metal object that can be used as a lever to open jars.
 - Kind of incline plane that reduces the force needed to cut.

58. Which of the following instructions for doing an experiment tells most clearly what to do and what to observe:
- A. Add 5 ml of sodium hydroxide to 50 ml of grape juice.
 - B. Add sodium hydroxide to grape juice and the juice will change color.
 - C. The hydroxide **concentration** of some substances is indicated by their color. . . .
 - D. Grape juice contains colored indicators.

Use the following contour map to answer questions 59 to 61.

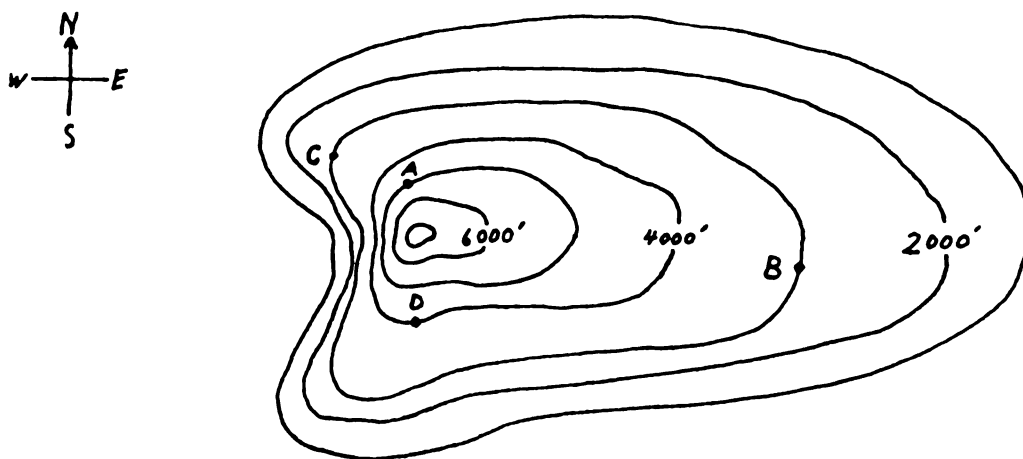


Figure 16

59. What is the elevation at point A:
- A. 7000 feet.
 - B. 6000 feet.
 - C. 5000 feet.
 - D. 4000 feet.
60. This mountain is steepest on its:
- A. North side.
 - B. South side.
 - C. East side.
 - D. West side.
61. Which of the following is at the highest elevation:
- A. A.
 - B. B.
 - C. C.
 - D. D.

APPENDIX IV - E
ITEM ANALYSIS FORM C

APPENDIX IV - E

ITEM ANALYSIS FORM C

All values are in Percent

Normal: Traditional item analysis procedure is used.

ICM: External criterion referenced item analysis as described in Chapter I

Item numbers in parentheses related to form A items.

Item	Key	Alt.	NORMAL			Diff. Disc.	ICM			Diff. Disc. Corr.
			U27	M46	L27		U27	M46	L27	
1 (1)	*	A	0	17	14	44	7	4	29	44
		B	79	46	50	29	64	63	36	29
		C	21	38	36		29	33	36	15
		D	0	0	0		0	0	0	
2 (2)	*	A	57	25	29	65	50	25	36	65
		B	7	21	14	29	14	17	14	14
		C	21	38	43		14	46	36	15
		D	14	17	14		21	13	14	
3 (3)		A	0	0	7	31	0	4	0	31
		B	7	25	43	43	14	8	64	50
		C	7	0	7		7	0	7	45
		* D	86	75	43		79	88	29	
4 (4)		A	7	13	21	48	7	13	21	48
		B	0	17	36	64	0	8	50	57
		C	14	12	29		14	25	7	38
		* D	79	58	14		79	54	21	
5 (5)	*	A	36	42	36	56	43	33	43	56
		B	57	46	29	39	50	50	29	21
		C	7	13	21		7	13	21	22
		D	0	0	14		0	4	7	
6 (6)	*	A	7	25	50	67	7	38	29	67
		B	57	29	14	43	50	29	21	29
		C	21	29	29		21	29	29	26
		D	14	17	7		21	4	21	

Appendix IV-E cont'd

Item	Key	Alt.	NORMAL			Diff. Disc.	ICM			Diff. Disc. Corr.
			U27	M46	L27		U27	M46	L27	
7 (7)		A	0	13	14	62	0	21	0	62
		B	14	8	21	14	21	8	14	-7
		C	36	46	29		36	42	36	0
	*	D	50	33	36		43	29	50	
8 (8)		A	57	42	29	81	57	36	81	
		B	7	25	14	14	7	21	21	14
		C	7	17	43		7	25	29	08
	*	D	29	17	14		29	17	14	
9 (10)		A	43	21	21	63	43	21	21	63
		B	0	13	43	43	0	21	29	36
	*	C	50	46	7		50	42	14	33
		D	7	21	29		7	17	36	
10 (11)		A	0	0	14	50	0	4	7	50
		B	0	8	0	64	0	4	7	50
	*	C	79	54	14		86	38	36	
		D	21	38	71		14	54	50	
11 (12)		A	29	63	57	63	36	58	57	63
	*	B	64	33	14	50	57	29	29	29
		C	0	0	29		0	8	14	41
		D	7	4	0		7	4	0	
12 (13)		A	14	21	7	60	14	17	14	60
		B	21	29	43	14	29	25	43	14
	*	C	50	38	36		50	38	36	14
		D	14	13	14		7	21	7	
13 (14)		A	7	25	29	37	7	29	21	37
		B	7	13	7	36	14	8	7	14
		C	0	4	14		0	8	7	26
	*	D	86	58	50		79	54	64	
14 (16)		A	7	13	14	65	7	8	21	65
		B	21	29	57	57	21	42	36	43
	*	C	64	33	7		57	33	14	37
		D	7	25	21		14	17	29	

Appendix IV-E cont'd

Item	Key	Alt.	NORMAL			Diff. Disc.	ICM			Diff. Disc. Corr.
			U27	M46	L27		U27	M46	L27	
15 (17)		A	21	33	29	79	29	29	29	79
		B	0	0	14	29	0	0	14	14
		C	43	46	50		36	58	36	11
		D	36	21	7		36	13	21	
16 (18)	*	A	93	75	71	21	100	71	71	21
		B	0	17	14	21	0	13	21	29
		C	0	0	0		0	0	0	26
		D	7	8	14		0	17	7	
17 (19)		A	7	29	7	48	7	25	14	48
		B	43	17	14	0	50	21	0	-21
	*	C	50	54	50		43	50	64	-06
		D	0	0	29		0	4	21	
18 (20)	*	A	7	50	36	58	0	46	50	58
		B	79	33	21	57	86	29	21	64
		C	14	8	36		14	13	29	49
		D	0	8	7		0	13	0	
19 (21)	*	A	7	17	36	44	0	21	36	44
		B	86	54	29	57	93	50	29	64
		C	7	25	14		7	25	14	45
		D	0	0	21		0	0	21	
20 (23)		A	0	8	21	50	0	4	29	50
		B	0	0	7	36	0	4	0	29
	*	C	71	46	36		71	42	43	10
		D	29	46	36		29	50	29	
21 (32)		A	14	21	43	58	14	17	50	58
		B	0	13	29	71	7	8	29	79
	*	C	79	42	7		79	46	0	54
		D	7	25	21		0	29	21	
22 (33)	*	A	7	13	21	38	7	13	21	38
		B	93	58	36	57	93	54	43	50
		C	0	13	36		0	17	29	39
		D	0	17	7		0	17	7	

Appendix IV-E cont'd

Item	Key	Alt.	NORMAL			Diff. Disc.	ICM			Diff. Disc. Corr.
			U27	M46	L27		U27	M46	L27	
23 (35)	*	A	7	25	50	67	14	25	43	67
		B	64	21	21	43	57	21	29	29
		C	14	25	14		14	21	21	34
		D	14	29	14		14	33	7	
24 (36)	*	A	21	25	21	71	36	21	14	71
		B	36	29	21	0	29	29	29	-7
		C	29	29	29		29	25	36	-11
		D	14	17	29		7	25	21	
25 (37)	*	A	21	33	14	56	21	25	29	56
		B	7	4	36	29	7	13	21	43
		C	64	38	36		64	46	21	32
		D	7	25	14		7	17	29	
26 (38)	*	A	43	54	50	67	50	63	29	67
		B	36	38	21	14	29	29	43	-14
		C	21	8	29		21	8	29	-06
		D	0	0	0		0	0	0	
27 (39)	*	A	93	79	21	33	86	75	36	33
		B	0	13	50	71	7	13	43	50
		C	7	8	21		7	13	14	44
		D	0	0	7		0	0	7	
28 (40)	*	A	7	33	29	52	21	25	29	52
		B	64	46	36	29	64	42	43	21
		C	29	21	36		14	33	29	15
		D	0	0	0		0	0	0	
29 (41)	*	A	50	75	57	83	57	75	50	83
		B	14	21	21	14	14	17	29	7
		C	36	4	21		29	8	21	09
		D	0	0	0		0	0	0	
30 (42)	*	A	79	29	7	63	71	33	7	63
		B	7	13	29	71	21	17	7	64
		C	7	33	43		0	33	50	
		D	7	21	21		7	13	36	

Appendix IV-E cont'd

Item	Key	Alt.	NORMAL			Diff. Disc.	ICM			Diff. Disc. Corr.
			U27	M46	L27		U27	M46	L27	
31 (43)	*	A	7	21	43	65	7	29	29	65
		B	57	33	14	43	57	33	14	43
		C	7	21	21		7	13	36	25
		D	29	25	21		29	25	21	
32 (44)		A	7	38	36	62	7	33	43	62
		B	0	17	36	64	0	17	36	71
		C	14	17	14		14	17	14	44
		D	79	29	14		79	33	7	
33 (45)	*	A	86	46	21	50	86	50	14	50
		B	7	17	14	64	14	13	14	71
		C	0	8	21		0	13	14	48
		C	7	29	43		0	25	57	
34 (46)	*	A	14	38	43	40	21	25	57	40
		B	86	63	29	57	79	67	29	50
		C	0	0	21		0	8	7	
		D	0	0	7		0	0	7	
35 (47)	*	A	7	4	29	63	7	8	21	63
		B	14	58	29	43	14	58	29	29
		C	64	29	21		64	21	36	25
		D	14	8	21		14	13	14	
36 (48)	*	A	86	50	36	75	86	58	21	75
		B	7	25	43	-36	7	17	57	-50
		C	7	17	14		7	21	7	-47
		D	0	8	7		0	4	14	
37 (49)		A	0	4	29	79	0	8	21	79
		B	0	25	29	43	0	29	21	21
		C	57	50	43		64	46	43	21
		D	43	21	0		36	17	14	
38 (51)	*	A	64	75	14	44	64	63	36	44
		B	0	4	21	50	0	13	7	29
		C	0	4	50		0	17	29	26
		D	36	17	14		36	8	29	

Appendix IV-E cont'd

Item	Key	Alt.	NORMAL			Diff. Disc.	ICM			Diff. Disc. Corr.
			U27	M46	L27		U27	M46	L27	
39 (52)	*	A	0	21	21	56	0	17	29	56
		B	21	8	14	36	14	13	14	36
		C	57	50	21		64	42	29	30
		D	21	21	43		21	29	29	
40 (53)	*	A	0	21	7	50	7	13	14	50
		B	57	50	43	14	57	50	43	14
		C	0	8	36		0	13	29	08
		D	43	21	14		36	25	14	
41 (55)	*	A	0	8	21	21	7	8	14	21
		B	0	8	0	50	7	4	0	21
		C	100	83	50		86	83	64	24
		D	0	0	29		0	4	21	
42 (57)	*	A	0	4	21	56	0	8	14	56
		B	7	13	21	21	14	8	21	0
		C	57	42	36		50	38	50	0
		D	36	42	21		36	46	14	
43 (58)	*	A	100	83	36	25	100	79	43	25
		B	0	0	29	64	0	8	14	57
		C	0	4	0		0	0	7	53
		D	0	13	36		0	13	36	
44 (59)	*	A	0	0	7	63	0	4	0	63
		B	36	33	43	-7	29	38	43	7
		C	29	42	36		36	42	29	06
		D	36	25	14		36	17	29	
45 (60)	*	A	0	4	29	63	0	8	21	63
		B	43	42	21	29	36	38	36	36
		C	7	13	29	7	7	17	21	36
		D	50	38	21		57	33	21	
46 (61)	*	A	7	21	21	79	0	29	14	79
		B	29	13	29	0	43	8	21	21
		C	21	8	29		21	17	14	13
		D	43	58	21		36	46	50	

Appendix IV-E cont'd

Item	Key	Alt.	NORMAL			Diff. Disc.	ICM			Diff. Disc. Corr.
			U27	M46	L27		U27	M46	L27	
47 (62)	*	A	43	38	29	63	43	33	36	63
		B	14	17	50	14	29	13	43	7
		C	14	4	7		7	13	0	01
		D	29	42	14		21	42	21	
48 (63)		A	7	21	7	48	7	21	7	48
		B	7	29	14	7	7	21	29	21
		C	21	8	21		21	8	21	10
		* D	64	42	57		64	50	43	
49 (64)		A	0	4	21	33	0	8	14	33
		B	0	4	21	79	0	8	14	79
		C	0	17	36		0	17	36	52
		* D	100	75	21		100	67	36	
50 (65)		A	7	8	57	44	0	17	50	44
		B	0	13	7	71	0	8	14	21
		* C	86	63	14		100	50	21	56
		D	7	17	21		0	25	14	
51 (66)		A	0	8	36	52	0	21	14	52
		B	0	13	14	29	0	13	14	21
		* C	57	54	29		64	42	43	08
		D	43	21	21		36	21	29	
52 (68)	*	A	21	29	14	77	29	17	29	77
		B	14	13	43	7	7	21	36	0
		C	14	21	36		14	29	21	
		D	50	33	7		50	29	14	
53 (69)		A	0	8	7	56	0	4	14	56
		B	36	25	50	43	29	29	50	50
		C	0	21	14		0	25	7	27
		* D	64	46	21		71	42	21	
54 (70)		A	0	13	36	54	0	17	29	54
		B	0	13	0	64	0	13	0	79
		C	21	8	14		7	8	29	57
		* D	79	46	14		93	38	14	

Appendix IV-E cont'd

Item	Key	Alt.	NORMAL			Diff. Disc.	ICM			Diff. Disc. Corr.
			U27	M46	L27		U27	M46	L27	
55 (71)	*	A	50	38	29	62	50	29	43	62
		B	0	13	29	21	0	17	21	7
		C	36	29	36		29	42	21	02
		D	7	13	0		14	8	0	
56 (73)	*	A	0	8	7	54	7	8	0	54
		B	0	8	21	86	0	8	21	79
		C	93	42	7		93	38	14	53
		D	0	25	21		0	17	36	
57 (76)	*	A	0	8	43	56	0	8	43	56
		B	0	33	14	64	0	29	21	64
		C	79	42	14		86	33	21	38
		D	21	13	21		14	25	7	
58 (74)	*	A	14	17	36	65	14	21	29	65
		B	64	25	21	43	57	25	29	29
		C	14	42	43		21	38	43	23
		D	7	17	0		7	17	0	
59 (77)	*	A	0	13	29	50	7	17	14	50
		B	7	21	36	50	14	21	29	43
		C	79	46	29		64	58	21	38
		D	14	21	7		14	4	36	
60 (78)	*	A	21	13	21	58	21	17	14	58
		B	0	0	7	43	0	4	0	36
		C	14	42	50		14	38	57	31
		D	64	42	21		64	38	29	
61 (79)	*	A	79	50	43	44	64	58	43	44
		B	0	13	7	36	0	13	7	21
		C	14	25	21		29	17	21	26
		D	7	8	21		7	8	21	

APPENDIX IV-F

VALIDATION SAMPLE SCORES

Student Number	TSPT SCORES		Individual Competency Measures					SRA	
	Form C	Form D	ID	CV	FH	DO	Total	Science	Reading
1	42	30	22	41	23	13	95	38	59
2	26	18	21	40	23	13	69	32	47
3	38	27	20	40	22	13	89	36	56
4	21	12	20	38	22	13	55	8	19
5	26	20	20	38	22	12	57	34	46
6	17	8	20	38	21	11	44	19	29
7	23	8	19	37	21	11	50	26	40
8	33	20	19	37	21	11	77	32	56
9	17	10	19	36	21	11	33	10	14
10	37	27	19	36	21	10	94	34	56
11	33	22	18	36	21	10	86	28	47
12	22	14	18	35	20	10	43	28	32
13	30	22	18	35	20	10	75	35	54
14	17	6	18	34	20	9	55	7	18
15	20	14	18	34	19	9	61	21	44
16	20	17	18	33	19	9	61	26	32
17	47	28	17	33	19	9	94	33	58
18	32	20	17	33	19	9	79	34	54
19	16	8	17	33	19	8	42	11	21
20	50	34	17	33	19	8	86	37	58
21	29	18	17	32	18	8	62	30	36
22	19	12	17	32	18	8	49	23	29
23	26	16	17	31	18	8	64	31	42
24	29	19	16	30	18	8	68	34	49
25	31	22	16	30	18	7	75	28	52
26	23	8	16	30	18	7	42	26	35
27	33	23	16	30	17	7	75	29	43
28	25	16	15	29	17	7	83	30	52
29	39	23	15	28	17	7	84	35	57
30	13	8	15	28	17	7	71	28	36
31	32	22	15	27	17	7	75		47
32	38	29	14	25	17	7	93		50
33	25	15	14	25	16	6	50	22	35
34	15	5	14	25	16	6	66	18	20
35	27	15	14	25	15	6	68	30	54
36	23	16	13	24	15	6	59	26	51
37	28	17	13	23	15	6	57	26	50
38	11	5	13	22	14	6	47	16	29
39	46	31	12	22	13	6	88	33	56
40	22	12	12	21	13	6	53	16	21

Appendix IV-F cont'd

<u>Student Number</u>	<u>TSPT SCORES</u>		<u>Individual Competency Measures</u>					<u>SRA</u>	
	<u>Form C</u>	<u>Form D</u>	<u>ID</u>	<u>CV</u>	<u>FH</u>	<u>DO</u>	<u>Total</u>	<u>Science</u>	<u>Reading</u>
41	12	7	11	21	13	5	66	22	36
42	31	18	10	21	13	5	75	25	45
43	23	17	10	20	12	5	79	21	35
44	16	9	10	19	12	5	47	23	35
45	18	4	10	18	12	4	33	14	22
46	12	4	10	18	12	4	57	16	23
47	38	24	9	18	12	3	86	39	54
48	29	17	9	18	12	3	58	18	25
49	39	28	9	17	11	2	83	37	53
50	36	26	8	16	10	2	80	32	44
51	24	10	7	16	9	2	67	31	42
52	28	21	6	9	8	2	78	29	52

APPENDIX IV-G

TSPT FORM C ITEMS - INDIVIDUAL COMPETENCY MEASURES SUBTEST CORRELATIONS

Item	Item Subtest Correlations				Significant Differences <u>Alpha = 0.1 or *Alpha = 0.01</u>
	ID	CV	FH	DO	
1	21	13	<u>13</u>	0	ID-DO
2	13	14	<u>12</u>	14	
3	<u>38</u>	42	<u>45</u>	25	CV-DO, FH-DO
4	<u>41</u>	30	40	27	
5	<u>18</u>	20	21	23	
6	21	25	18	<u>24</u>	
7	0	-04	<u>02</u>	-01	
8	<u>07</u>	09	<u>02</u>	01	
9	<u>28</u>	33	<u>28</u>	16	CV-DO
10	30	32	<u>31</u>	<u>37</u>	
11	<u>39</u>	38	41	<u>28</u>	
12	<u>12</u>	13	16	13	
13	<u>33</u>	22	20	22	
14	40	35	<u>26</u>	26	ID-FH, ID-DO
15	09	16	<u>04</u>	02	
16	20	18	<u>30</u>	33	DO-CV
17	-06	-02	-08	-12	
18	42	46	40	<u>36</u>	
19	34	42	36	<u>45</u>	
20	08	12	<u>03</u>	<u>08</u>	
21	52	<u>40</u>	<u>52</u>	50	ID-CV, FH-CV
22	45	<u>32</u>	37	21	ID-CV, ID-DO, FH-DO
23	22	<u>28</u>	38	39	FH-ID, DO-ID
24	-06	-15	-15	08	DO-CV, DO-FH
25	16	<u>35</u>	24	32	CV-ID, DO-ID
26	<u>02</u>	-07	-08	-05	
27	<u>47</u>	41	43	21	ID-DO*, CV-DO, FH-DO
28	<u>03</u>	09	19	23	FH-ID, DO-ID
29	<u>07</u>	09	05	12	
30	<u>42</u>	44	40	37	
31	33	13	26	<u>27</u>	ID-CV, FH-CV
32	38	38	45	<u>26</u>	FH-DO
33	38	42	47	<u>43</u>	
34	<u>36</u>	45	37	<u>16</u>	ID-DO, CV-DO*, FH-DO
35	<u>19</u>	21	17	32	
36	-39	-42	-51	-30	DO-FH
37	<u>10</u>	18	23	26	DO-ID
38	28	29	22	<u>13</u>	ID-DO, CV-DO
39	22	<u>31</u>	22	<u>23</u>	
40	20	<u>04</u>	<u>01</u>	02	ID-CV, ID-FH, ID-DO
41	17	29	<u>16</u>	05	CV-FH, CV-DO
42	02	03	-05	-02	
43	53	<u>51</u>	40	<u>36</u>	ID-FH, ID-DO, CV-DO
44	0	<u>04</u>	14	<u>06</u>	FH-ID
45	29	<u>32</u>	38	<u>22</u>	FH-DO

Appendix IV-G cont'd

Item	Item Subtest Correlations				Significant Differences Alpha = 0.1 or *Alpha = 0.01
	ID	CV	FH	DO	
46	04	09	18	<u>13</u>	FH-ID
47	06	-03	06	<u>0</u>	
48	07	11	04	<u>16</u>	
49	43	50	<u>42</u>	<u>37</u>	
50	38	52	50	<u>56</u>	CV-ID, DO-ID
51	19	<u>05</u>	0	<u>14</u>	ID-CV, ID-FH
52	-02	<u>-07</u>	-03	-05	
53	<u>40</u>	<u>21</u>	11	34	ID-CV, ID-FH*, DO-FH
54	<u>45</u>	54	38	60	DO-ID, CV-FH, DO-FH
55	<u>03</u>	04	-15	22	ID-FH, CV-FH, DO-CV, DO-FH*
56	<u>44</u>	56	38	<u>38</u>	CV-ID, CV-FH, CV-DO
57	41	33	28	<u>27</u>	ID-FH
58	19	23	17	<u>16</u>	
59	<u>28</u>	34	45	20	FH-ID, FH-DO
60	<u>18</u>	29	39	19	FH-ID, FH-DO
61	<u>15</u>	20	41	11	FH-ID*, FH-CV, FH-DO*

ID = Interpreting Data subtest of the Individual Competency Measures

CV = Controlling Variables subtest of the Individual Competency Measures

FH = Formulating Hypotheses subtest of the Individual Competency Measures

DO = Defining Operationally subtest of the Individual Competency Measures

All decimal points are suppressed.

The underlined correlation indicates the subtest which the item is intended to assess.

Correlations which are significantly different from each other are indicated in the last column. The level of confidence is 0.1 except for the differences followed by an asterisk, which have a confidence level of 0.01.

This table may be interpreted as follows:

Item 1: The highest correlation is with ID, but since the correlations with CV and FH are not significantly lower, it cannot be said that this item assesses any one of them uniquely. Since the difference between either the CV or FH correlation and the DO correlation is not significant either, it cannot be said that if the item measures CV or FH it does not measure DO. All that can be said with 90 percent confidence is that if it measures ID, it does not measure DO.

Appendix IV-G cont'd

Item 22: The highest correlation is with ID, but the difference between the ID and the FH correlations is not significant so it is tempting to say this item measures either ID or FH. However, the difference between the FH and CV correlations is not significant so if the item measures FH, it could be measuring CV too. But the difference between CV and DO is not significant either so if CV is allowed, DO cannot be excluded. Thus although this item is not as ambiguous as number one above, it is only safe to say that it probably does not measure DO.

Item 40: The highest correlation is with ID and the difference between the ID correlation and any of the other correlations is significant at the 0.1 level. Thus it can be said with 90 percent confidence that students' scores on this item indicate their ability to use the process of Interpreting Data more than any of the other Integrated Processes. Therefore, it is safe to label this item as an "ID item." How this interpretation can be harmonized with a logical analysis of the item is not of concern here.

APPENDIX IV-H

TSPT FORM D

THE
SCIENCE
PROCESSES
TEST

FORM D
(1974)

by

ROBERT R. LUDEMAN
DARRELL W. FYFFE
RICHARD W. ROBISON
RICHARD J. MCLEOD
GLENN D. BERKHEIMER

COPYRIGHT BY
ROBERT R. LUDEMAN
1974

USE THE ANSWER SHEET PROVIDED
PLEASE DO NOT MAKE ANY MARKS ON THIS BOOKLET

The graph on the right was made by setting weights on the end of the stick which is clamped to the table as shown in Figure 1.

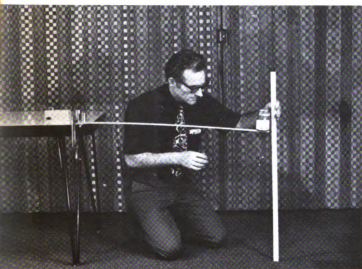


Figure 1

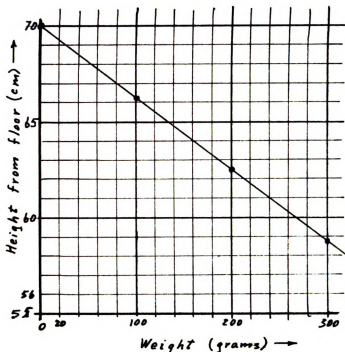


Figure 2

1. The can of soup shown above bends the stick to 66 cm from the floor. The can weighs about:

- A. 67.5 grams
- B. 110 grams
- C. 255 grams
- D. 315 grams

Questions 2, 3 and 4 use this set up:

Several identical sticks are stacked on top of each other and extended over the edge of a table in such a way as to give the greatest possible overhang (see below).

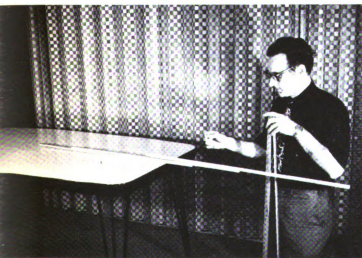


Figure 3

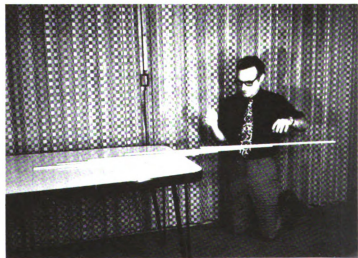


Figure 4

The relation between greatest overhang and number of sticks is graphed below:

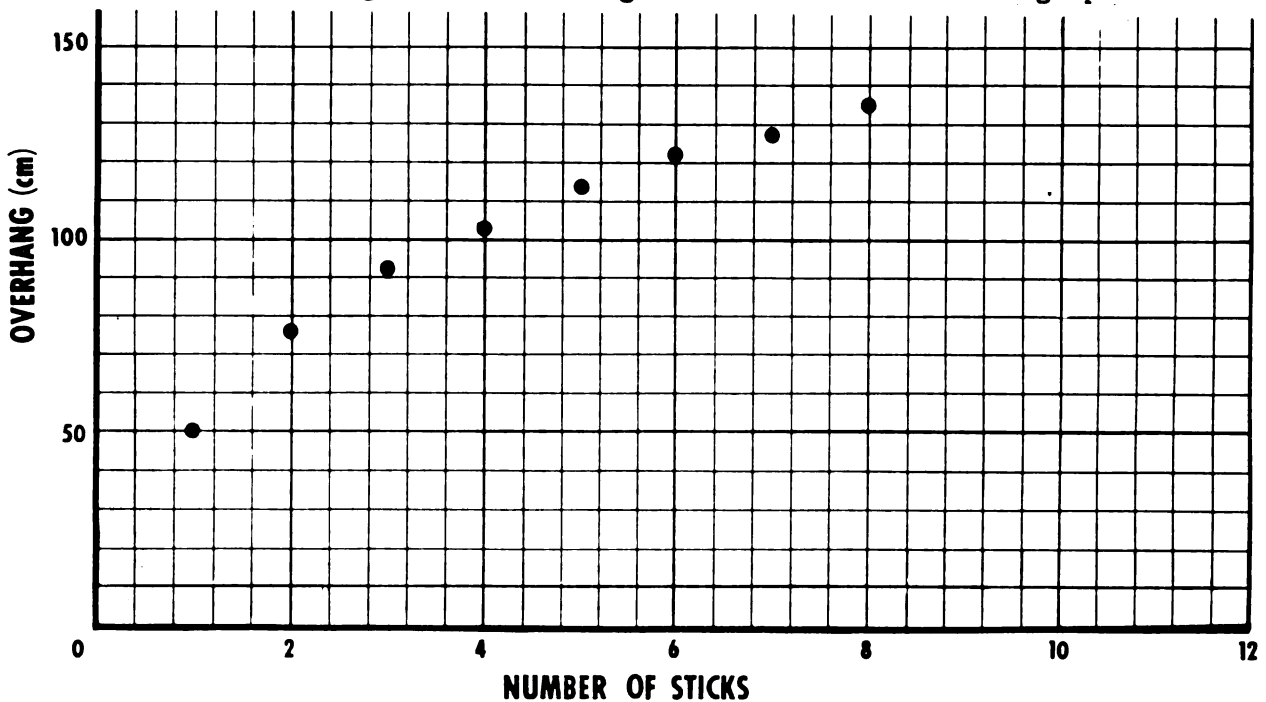


Figure 5

2. The greatest overhang you could get using 5 sticks would be about:
 - A. 1 cm.
 - B. 49.9 cm.
 - C. 99 cm.
 - D. 112 cm.
3. The smallest number of sticks you would need to get an overhang of 100 cm is:
 - A. 1 stick
 - B. 2 sticks
 - C. 3 sticks
 - D. 4 sticks
4. Using 10 sticks you could get a maximum overhang of:
 - A. More than 150 cm.
 - B. Between 130 and 150 cm.
 - C. Between 110 and 130 cm.
 - D. Less than 110 cm.
5. If you needed to tell someone what I mean above by "overhang" so that they could measure the overhang, you could say "overhang" means the distance from the:
 - A. Far end of the top stick to the center of gravity of the system.
 - B. Far end of the top stick to the edge of the table.
 - C. Far end of the top stick to the far end of the next stick.
 - D. Center of gravity of the top stick to the edge of the table.

Questions 6, 7 and 8 are about this experiment:



Figure 6

The two jars shown are filled full to the brim.
The lids are screwed on tight.



Figure 7

The jars are put in a pan and placed in a freezer.
The temperature inside the freezer is 0 degrees F.

6. Two hours later it is found that neither of the liquids in the jars is frozen. John says the reason is that two hours is not long enough to freeze water. Tom says the reason is that the liquid in the jars is not water, but is some kind of antifreeze.
- A. Probably John is right and Tom is wrong.
 - B. Probably John is wrong and Tom is right.
 - C. Both John and Tom could be right.
 - D. It is unreasonable that either John or Tom is right.
7. Next day when the freezer is opened, jar "Y" is found broken and its contents are frozen solid: The contents of jar "X" is still liquid. You now know that:
- A. Jar "X" and "Y" do not contain the same kind of liquid.
 - B. At least one of the jars contains water.
 - C. The temperature of the jars is different.
 - D. All of the above are correct.
8. Suppose someone told you that the contents of jar "Y" behaved like a "Cronon" while the contents of jar "X" behaved like a "non Cronon." To show the difference between them you can say a "Cronon" is:
- A. A chemical, but a non Cronon is not.
 - B. Definitely not water.
 - C. Easier to freeze than a non Cronon.
 - D. A special kind of jar.

9. A jar of water was left in the above freezer over night. The next day the jar is broken and the water is frozen. You know that:
- A. Water freezes easier than a Cronon.
 - B. Water behaves like a Cronon.
 - C. Cronons are made of water.
 - D. The water was cold before it was put in the freezer.
10. A bottle of alcohol was left in the above freezer over night. It was not frozen. You now know that:
- A. Alcohol is a Cronon.
 - B. Alcohol is a non Cronon.
 - C. Non Cronons are made of alcohol.
 - D. The alcohol was warm before it was put in the freezer.

Questions 11 through 19 are about the following experiment:

Cylinders of various shapes are rolled down the sloping table shown below.



Figure 8

The slope is kept constant by the books under the table legs. The "rolling time" is measured as the time it takes for the cylinders to roll the length of the table. Whether the cylinders are metal or plastic, short or long, solid or hollow is shown on the chart on the next page.

The Cylinders shown on the right are described as follows:

Metal cylinders: A, B, C, D.

Plastic cylinders: E, F, G, H.

Short cylinders: A, B, E, F.

Long cylinders: C, D, G, H.

Solid cylinders: A, C, E, G.

Hollow cylinders: B, D, F, H.



Figure 9

11. By comparing rolling times for cylinders A and E, you could test the effect of the variable:
 - A. Metal or plastic.
 - B. Short or long.
 - C. Solid or hollow.
 - D. Amount of slope.
12. By comparing rolling times for cylinders F and H you could test the effect of the variable:
 - A. Metal or plastic.
 - B. Short or long.
 - C. Solid or hollow.
 - D. Amount of slope.
13. It was found that the rolling time was exactly the same for cylinders B and D. From this information alone, the variable that does not affect the rolling time is:
 - A. Metal or plastic.
 - B. Short or long.
 - C. Solid or hollow.
 - D. Amount of slope.
14. Which cylinders could you use to tell if a metal cylinder rolls at a different rate than a plastic cylinder?
 - A. A and F.
 - B. A and H.
 - C. C and G.
 - D. C and H.

Rolling time for the above cylinders is given in the following table. Use it to answer questions 15 through 18.

<u>Cylinder</u>	<u>Material</u>	<u>Length</u>	<u>Type</u>	<u>Time</u>
A	metal	2 cm	solid	5 sec.
B	metal	2	hollow	10
C	metal	8	solid	5
D	metal	8	hollow	10
E	plastic	2	solid	5
F	plastic	2	hollow	10
G	plastic	8	solid	5
H	plastic	8	hollow	10

15. The variable solid or hollow affects the rolling time.
- A. True.
 - B. False.
 - C. Cannot tell from the data
16. The above table shows that for this experiment the rolling time for hollow cylinders:
- A. Is always the same.
 - B. Depends on how long the cylinders are.
 - C. Is different for metal than for plastic.
 - D. Depends on the slope of the table.
17. For this experiment a solid cylinder is one which:
- A. Is metal.
 - B. Has a rolling time of 5 seconds.
 - C. More than one of the above is correct.
 - D. None of the above is correct.
18. For this experiment a metal cylinder is one which:
- A. Is solid.
 - B. Has a rolling time of 5 seconds.
 - C. More than one of the above is correct.
 - D. None of the above is correct.
19. To tell someone how to **answer** the above questions it would be best to tell them that by "rolling time" I mean the time:
- A. For the cylinder to roll the length of the table.
 - B. During which gravity is pulling on the cylinder.
 - C. As shown by my stop watch.
 - D. Needed for me to start and stop my stop watch.

Questions 20 through 24 are about the following experiment:

The TV ads say that their falseteeth cleaner colors the water green. When the color goes away, your teeth are clean. To find the time needed for the color to go away I got several glasses of water. Each at a different temperature. I put a tablet of falseteeth cleaner in each glass and measured the time needed for the green color to go away. A graph of the data is shown below.

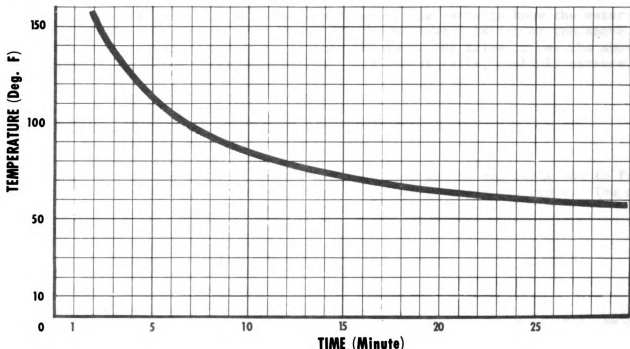


Figure 10

20. From this graph we can say that for a higher temperature, the time needed for the green color to go away is:
- A. Greater.
 - B. Less.
 - C. Not affected by changes in temperature.
 - D. Not enough information is given to tell.
21. From the graph, how long should it take to clean your falseteeth at a temperature of 70 degrees F.
- A. 5 minutes or less.
 - B. Close to 10 minutes.
 - C. Close to 15 minutes.
 - D. 20 minutes or more.
22. If you needed to clean your falseteeth in 5 minutes, you should use a temperature of:
- A. Less than 60 degrees.
 - B. Between 60 and 75 degrees.
 - C. Between 75 and 100 degrees.
 - D. Greater than 100 degrees.

23. To tell a friend how to measure the cleaning time it would be best to say the cleaning time is the time for:
- A. The green color to go away.
 - B. The chemical reaction to be completed.
 - C. All the bacteria on the teeth to be killed.
 - D. All the above are equally good answers.
24. Suppose a friend has gone to a lake to swim. He wants to know the water temperature but he has no thermometer. He borrows a tablet of the above falseteeth cleaner, and drops it into the lake. He tries to use the above graph to tell the water temperature. His effort fails, probably because:
- A. The water is too cold to swim in.
 - B. He did not wait long enough.
 - C. He used the wrong amount of water.
 - D. There are no falseteeth in the water.
25. Suppose a friend dials a number, hands you the phone, and tells you to find out if the store he has called is a hardware store or a grocery store. You can ask only one of the following questions. You should ask:
- A. If they sell can openers.
 - B. If they sell light bulbs.
 - C. What they sell the most of.
 - D. How old their store is.
26. A scientist wanted to know if a special light bulb is as "efficient" as sunlight. He selected two young bean plants. He placed one plant in his window and the other in a closet. He put his special light bulb in the socket in the closet, turned it on, and closed the door. He returned in three days to see how his plants were doing. He found that both plants had grown exactly the same amount. Therefore he decided his special light bulb is as "efficient" as sunlight. The reason the scientist used 2 plants in the experiment is:
- A. So he could compare the plants.
 - B. In case one plant died, the experiment would not be a failure.
 - C. He really did not need 2 plants.
 - D. His chances of getting a healthy plant were better by using two plants than if he had chosen only one.
27. All objects can be bent by some small amount no matter how stiff they are. This idea must be accepted until:
- A. No one believes it any more.
 - B. A scientist says it is no longer true.
 - C. Objects are found that bend easily.
 - D. Someone finds an object that does not bend.
28. Mary has a thermometer in her room. Her thermometer is best described as:
- A. An indoor thermometer.
 - B. A glass tube containing a colored liquid.
 - C. A device for measuring temperature.
 - D. A thermostat.

Questions 29, 30 and 31 are about this experiment:

A science class decided to check their reaction times. Each student was asked to flip a switch as soon as he saw a light flash, heard a buzzer sound, or both. A timer recorded the time it took for each student to react.

The class averages were figured. The results were:

STIMULUS	AVERAGE TIME	
	Boys	Girls
Light	.17 sec	.15 sec.
Buzzer	.22	.19
Both	.14	.23

29. Who reacted more quickly to the buzzer?

- A. Boys by .02 sec.
- B. Girls by .02 sec.
- C. Boys by .03 sec.
- D. Girls by .03 sec.

30. Who reacted more quickly to both the light and the buzzer together?

- A. Girls by .03 sec.
- B. Boys by .05 sec.
- C. Girls by .05 sec.
- D. Boys by .09 sec.

31. You are given a block of wood and a glass full of an unknown liquid. To find out whether the wood will float on the surface of the liquid you should:

- A. Find the density of the wood.
- B. Find the density of the liquid.
- C. Put the block of wood in the liquid and watch it.
- D. Put the block of wood in several different kinds of liquids and watch it.

32. A girl removed a lid from a jar by prying on it with the blade of a table knife. From that use of it, you might say a knife is a:

- A. Sterling silver object with a sharp edge and a decorated handle.
- B. Stainless steel object about 8 inches along with a thin blade.
- C. Metal object that can be used as a lever to open jars.
- D. Kind of incline plane that reduces the force needed to cut.

33. Which of the following instructions for doing an experiment tells most clearly what to do and what to observe:

- A. Add 5 ml of sodium hydroxide to 50 ml of grape juice.
- B. Add sodium hydroxide to grape juice and the juice will change color.
- C. The hydroxide concentration of some substances is indicated by their color.
- D. Grape juice contains colored indicators.

Use the following contour map to answer questions 34 through 36.

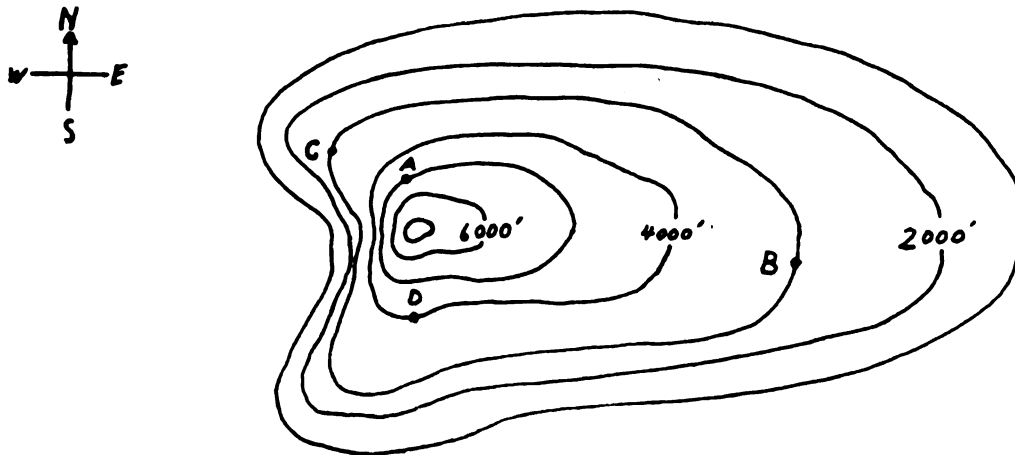


Figure 11

34. What is the elevation at point A:

- A. 7000 feet.
- B. 6000 feet.
- C. 5000 feet.
- D. 4000 feet.

35. This mountain is steepest on its:

- A. North side.
- B. South side.
- C. East side.
- D. West side.

36. Which of the following is at the highest elevation:

- A. A
- B. B
- C. C
- D. D

BLACKEN THE LETTER BOX BELOW WHICH MATCHES EACH LETTER OF YOUR NAME.		YOUR LAST NAME		YOUR FIRST NAME		MI
A	A	A	A	A	A	A
B	B	B	B	B	B	B
C	C	C	C	C	C	C
D	D	D	D	D	D	D
E	E	E	E	E	E	E
F	F	F	F	F	F	F
G	G	G	G	G	G	G
H	H	H	H	H	H	H
I	I	I	I	I	I	I
J	J	J	J	J	J	J
K	K	K	K	K	K	K
L	L	L	L	L	L	L
M	M	M	M	M	M	M
N	N	N	N	N	N	N
O	O	O	O	O	O	O
P	P	P	P	P	P	P
Q	Q	Q	Q	Q	Q	Q
R	R	R	R	R	R	R
S	S	S	S	S	S	S
T	T	T	T	T	T	T
U	U	U	U	U	U	U
V	V	V	V	V	V	V
W	W	W	W	W	W	W
X	X	X	X	X	X	X
Y	Y	Y	Y	Y	Y	Y
Z	Z	Z	Z	Z	Z	Z

THE
SCIENCE
PROCESSES
TEST

FORM D
(1974)

INSTRUCTIONS:

1. Print your name in the boxes above.
If you have a long name, there may not be enough space for all of your name. That will not matter.
2. For each question in the test booklet, black in all of the box below the letter on this answer sheet which is the best answer.
3. Notice that the question numbers go across the answer sheet. There is one line on this answer sheet for each page in the test booklet.

Page 1:	1	A B C D			
Page 2:	2	A B C D	3	4	5
Page 3:	6	A B C D	7	8	
Page 4:	9	A B C D	10		
Page 5:	11	A B C D	12	13	14
Page 6:	15	A B C D	16	17	18
Page 7:	20	A B C D	21	22	
Page 8:	23	A B C D	24	25	26
Page 9:	29	A B C D	30	31	32
Page 10:	34	A B C D	35	36	

BLACKEN THE LETTER BOX BELOW WHICH MATCHES EACH LETTER OF YOUR NAME.		YOUR LAST NAME		YOUR FIRST NAME		MI
A	A	A	A	A	A	A
B	B	B	B	B	B	B
C	C	C	C	C	C	C
D	D	D	D	D	D	D
E	E	E	E	E	E	E
F	F	F	F	F	F	F
G	G	G	G	G	G	G
H	H	H	H	H	H	H
I	I	I	I	I	I	I
J	J	J	J	J	J	J
K	K	K	K	K	K	K
L	L	L	L	L	L	L
M	M	M	M	M	M	M
N	N	N	N	N	N	N
O	O	O	O	O	O	O
P	P	P	P	P	P	P
Q	Q	Q	Q	Q	Q	Q
R	R	R	R	R	R	R
S	S	S	S	S	S	S
T	T	T	T	T	T	T
U	U	U	U	U	U	U
V	V	V	V	V	V	V
W	W	W	W	W	W	W
X	X	X	X	X	X	X
Y	Y	Y	Y	Y	Y	Y
Z	Z	Z	Z	Z	Z	Z

THE SCIENCE PROCESSES TEST

FORM D
(1974)

INSTRUCTIONS:

1. Print your name in the boxes above.
If you have a long name, there may not be enough space for all of your name. That will not matter.
2. For each question in the test booklet, black in all of the box below the letter on this answer sheet which is the best answer.
3. Notice that the question numbers go across the answer sheet. There is one line on this answer sheet for each page in the test booklet.

Page 1:	1	A B C D			
Page 2:	2	A B C D	3	A B C D	4
Page 3:	6	A B C D	7	A B C D	8
Page 4:	9	A B C D	10	A B C D	
Page 5:	11	A B C D	12	A B C D	13
Page 6:	15	A B C D	16	A B C D	17
Page 7:	20	A B C D	21	A B C D	22
Page 8:	23	A B C D	24	A B C D	25
Page 9:	29	A B C D	30	A B C D	31
Page 10:	34	A B C D	35	A B C D	36

APPENDIX IV-I
NORMING AREA MAP



APPENDIX IV - J
ITEM ANALYSIS FORM C

APPENDIX IV-J

ITEM ANALYSIS FORM D

Item number in parentheses refers to form C item

<u>Item</u>	<u>Key</u>	<u>Alt.</u>	<u>U27</u>	<u>M46</u>	<u>L27</u>	<u>Diff.</u>	<u>Disc.</u>
1 (11)	*	A	15	32	50	60	47
		B	69	34	22		
		C	7	11	15		
		D	9	23	13		
<hr/>							
2 (3)	*	A	1	7	12	39	58
		B	4	14	28		
		C	6	18	29		
		D	89	62	31		
<hr/>							
3 (4)	*	A	2	11	15	37	50
		B	2	10	18		
		C	7	18	28		
		D	89	52	39		
<hr/>							
4 (5)	*	A	39	37	34	57	30
		B	57	44	27		
		C	3	12	20		
		D	1	8	19		
<hr/>							
5 (6)	*	A	19	22	20	63	23
		B	51	34	28		
		C	20	25	29		
		D	10	20	23		
<hr/>							
6 (14)	*	A	9	14	18	57	24
		B	9	14	18		
		C	55	43	31		
		D	16	22	22		
<hr/>							
7 (16)	*	A	67	66	50	38	17
		B	7	13	19		
		C	3	10	17		
		D	24	11	14		

Appendix IV - J cont'd

<u>Item</u>	<u>Key</u>	<u>Alt.</u>	<u>U27</u>	<u>M46</u>	<u>L27</u>	<u>Diff.</u>	<u>Disc.</u>
8 (17)	*	A	18	22	27	54	44
		B	10	26	27		
		C	71	42	27		
		D	1	10	19		
<hr/>							
9 (18)	*	A	32	39	45	68	38
		B	54	28	16		
		C	12	5	17		
		D	2	28	22		
<hr/>							
10 (19)	*	A	16	25	33	50	49
		B	77	47	28		
		C	6	13	30		
		D	1	16	19		
<hr/>							
11 (21)	*	A	89	52	23	46	66
		B	1	17	25		
		C	5	18	30		
		D	6	14	21		
<hr/>							
12 (22)	*	A	2	10	18	40	54
		B	89	58	35		
		C	5	16	27		
		D	5	16	19		
<hr/>							
13 (23)	*	A	5	23	20	48	56
		B	86	45	30		
		C	8	18	28		
		D	7	15	22		
<hr/>							
14 (25)	*	A	14	21	28	59	50
		B	7	17	26		
		C	69	37	19		
		D	10	26	37		
<hr/>							
15 (27)	*	A	92	66	37		
		B	6	19	31		
		C	2	15	27		
		D	0	1	5		

Appendix IV - J cont'd

<u>Item</u>	<u>Key</u>	<u>Alt.</u>	<u>U27</u>	<u>M46</u>	<u>L27</u>	<u>Diff.</u>	<u>Disc.</u>
16 (30)	*	A	90	50	17	48	73
		B	3	5	29		
		C	7	18	29		
		D	1	27	26		
<hr/>							
17 (31)	*	A	4	18	32	51	49
		B	73	50	24		
		C	8	13	24		
		D	15	19	20		
<hr/>							
18 (32)		A	7	42	33	65	56
		B	8	18	27		
		C	13	18	23		
	*	D	73	23	17		
<hr/>							
19 (33)	*	A	91	57	30	41	61
		B	6	18	29		
		C	1	12	22		
		D	2	13	19		
<hr/>							
20 (34)	*	A	11	21	30	49	57
		B	86	43	29		
		C	0	11	21		
		D	3	26	20		
<hr/>							
21 (35)		A	0	11	25	44	64
		B	8	18	27		
	*	C	87	57	23		
		D	5	15	25		
<hr/>							
22 (37)		A	1	14	23	43	73
		B	1	16	31		
		C	4	14	24		
	*	94	56	21			
<hr/>							
23 (38)	*	A	39	38	19	67	20
		B	6	15	23		
		C	5	4	33		
		D	50	44	24		
<hr/>							
24 (39)		A	10	23	25	57	47
		B	6	15	24		
	*	C	69	40	22		
		D	15	22	29		

Appendix IV - J cont'd

Item	Key	Alt.	U27	M46	L27	Diff.	Disc.
25 (41)	*	A	4	5	18	34	51
		B	6	16	26		
		C	89	69	38		
		D	1	10	19		
<hr/>							
26 (43)	*	A	93	62	23	40	70
		B	2	12	21		
		C	2	16	29		
		D	3	11	27		
<hr/>							
27 (49)	*	A	1	17	18	42	58
		B	6	18	29		
		C	3	12	21		
		D	90	54	32		
<hr/>							
28 (50)	*	A	30	35	39	73	28
		B	3	16	19		
		C	44	23	16		
		D	23	27	25		
<hr/>							
29 (53)	*	A	1	10	18	53	54
		B	12	20	28		
		C	8	29	30		
		D	78	42	24		
<hr/>							
30 (54)	*	A	2	24	19	48	60
		B	1	13	24		
		C	9	19	29		
		D	88	45	28		
<hr/>							
31 (56)	*	A	5	12	18	44	55
		B	4	18	28		
		C	84	55	29		
		D	7	16	25		
<hr/>							
32 (57)	*	A	3	17	20	51	49
		B	10	28	27		
		C	76	46	27		
		D	11	9	26		

Appendix IV - J cont'd

<u>Item</u>	<u>Key</u>	<u>Alt.</u>	<u>U27</u>	<u>M46</u>	<u>L27</u>	<u>Diff.</u>	<u>Disc.</u>
33 (58)	*	A	21	33	21	61	41
		B	64	44	23		
		C	12	21	29		
		D	3	3	27		
<hr/>							
34 (59)	*	A	7	13	18	52	59
		B	6	35	41		
		C	83	42	24		
		D	4	11	17		
<hr/>							
35 (60)	*	A	4	12	19	51	51
		B	2	10	17		
		C	15	35	36		
		D	79	44	28		
<hr/>							
36 (61)	*	A	81	53	32	45	49
		B	3	10	17		
		C	14	25	36		
		D	1	12	14		

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 03145 9138