TABLE I

TEST OF SIGNIFICANCE OF DIFFERENCE IN GRADES IN GENERAL PHYSICS

Subgroups	K	X_2	X_5	X_6	Y_4	SS adjusted for X_2, X_5 and X_6 Total	Within	F (df: 1,109)
PSSC	57	26.877	22.702	20.123	2.825			
Conventional	57	26.316	24.070	19.000	2.614	79.59	78.68	1.26

K = Number of cases in each subgroup.
X_2 = Subgroup mean ACT score.
X_5 = Subgroup mean credit hours of college mathematics.
X_6 = Subgroup mean credit hours of other college physical science.
Y_4 = Subgroup mean average grade for sequence.

University between students who studied PSSC physics and students who studied conventional secondary school physics. A minimum value for F of 3.94 is necessary for significance at the five per cent level of confidence.

It was concluded on the basis of the above results that Hipsher's inference that PSSC students may be at a disadvantage with regard to success in college physics courses did not hold true at The Ohio State University. It is recommended that high school personnel not concern themselves with success in college physics when deciding whether to present PSSC or conventional physics courses and when counseling students as to which of the two courses to study.

THE DEVELOPMENT OF THE *TAB SCIENCE TEST* * †

DAVID P. BUTTS AND HOWARD L. JONES
Science Education Center, University of Texas, Austin, Texas 78712

INTRODUCTION

RECENT developments in science curricula have influenced a shift in emphasis for science instruction in today's elementary schools. Previously, instruction was directed toward helping the student learn a pre-selected body of scientific facts. Recently, however, much has been said concerning the merit of emphasizing the means by which these facts are obtained. Thus a major goal for science classes has become the development of student abilities necessary to carry on the processes of scientific inquiry.

As with all types of instruction, evaluation should be an integral part of curriculum development in science. Primarily, it is important that the instruction be evaluated by means of instruments which measure the competencies that are included within the goals and philosophy of the program. In content-oriented science classrooms the major goal is the child's learning of scientific facts and principles. Evaluation of this goal is possible through measuring the child's recall of factual information on a short-answer or essay test. In a curriculum oriented to include both the knowledge of science and the processes by which this knowledge is generated, evaluation must reflect not only what the student

* A paper presented at the thirty-ninth annual meeting of the National Association for Research in Science Teaching, February 19, 1966. Chicago, Illinois.

† The research reported in this paper is supported by a grant from this Cooperative Research Program of the Office of Education, U.S. Department of Health, Education, and Welfare.

knows, but also his ways of obtaining knowledge. However, tests which are designed by teachers for individual classes or current published tests which purport to measure students' inquiry methods provide little or no evaluation of the processes employed by different individuals to inquire into problem situations. In these tests, it is assumed that by knowing which answer the student selects, knowledge of the method that he used to solve the problem is available. Although it may appear that problem-solving processes can be inferred, studies usually show that students come up with ways at arriving at answers, often correct, that no teacher seems to have anticipated. (Bloom, 1956, p. 27)

In this paper, a procedure for the study of student inquiry behaviors will be described. This procedure analyzes the methods as well as the end products of inquiry by investigating the type, number, and sequence of questions asked by elementary school children when solving problems.

WHAT ARE INQUIRY BEHAVIORS?

In the analysis of recent changes in science education much attention has been given to the role of inquiry. However, there is a notable absence of an unambiguous description of the behaviors involved in inquiry. Most curriculum developments propose that, given a science problem to investigate, the student should be led to comprehend that there are some more productive ways of inquiry that might help him explore the problem. Suchman (1962) has analyzed in depth how inquiry is conducted by the elementary school child. He has related that in inquiring into a problem situation, a child (1) searches, (2) processes data, (3) discovers, and (4) verifies. However,

> while none of these actions is unique to inquiry, they are all essential to it, and in combination, form a cycle of operation that characterizes the inquiry process (Suchman, 1962, p. 5).

It has also been hypothesized that what is found through inquiry by the child often leads to the expansion of his conceptual systems through what Piaget (1950) calls assimilation and accommodation.

These five activities (searching, processing data, discovering, verifying, assimilating-accommodating) are the specifics for a model of inquiry behavior.

The Inquiry Cycle

The actions of inquiry are not separate and discrete entities, for each activity is dependent on others that have preceded or that will proceed. Perhaps a suitable model of inquiry behaviors would resemble that in Figure 1.

A child who finds himself in a problem situation may search for and process data. However, his searching and data processing operations are dependent on his past searching, data processing, verifying, discovering, assimilating, and accommodating behaviors or strategies. In turn, the child's verifying, discovering, assimilating, and accommodating processes are also dependent upon his searching and data processing behaviors.

Inquiry, then, may be represented as a cyclic operation. The end products of inquiry are the expansion of conceptual systems and the development of abilities to use these expanded systems to better inquire into other situations.

THE INSTRUMENT

If inquiry is learned, some children may be expected to be more proficient inquirers than others. However, the evidence of measurable differences in inquiry among children is questionable.

In the common essay or short-answer test, inquiry is measured by assuming that if the child knows the answer to a proposed problem, he is a proficient searcher, data processor, discoverer, verifier, assimilator, and accommodator. In the essay or short-answer test no indication is given of those behaviors the child employs in inquiring into a problem situation.

In this study when a child is to be tested as an inquirer, the evaluation of the be-

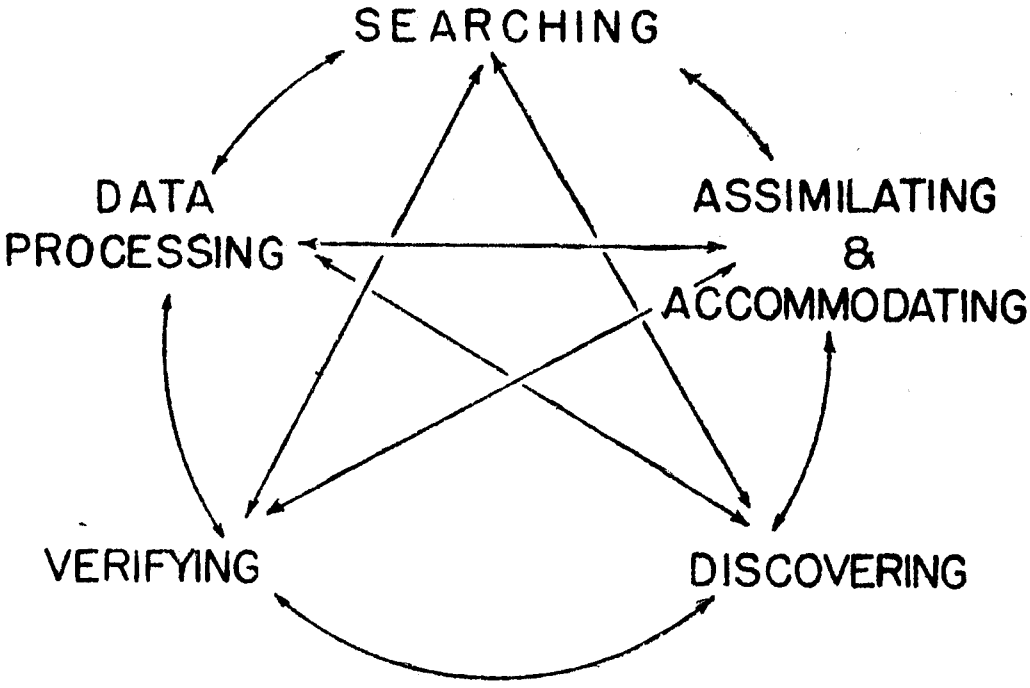

FIG. 1. A model of inquiry behavior.

haviors presented in the inquiry model are inferred from his performance. That is, the inquiry behaviors of a child are evaluated by analyzing his methods of (1) searching, (2) processing data, (3) discovering, (4) verifying. In addition, the child is allowed to apply conceptual understandings in new situations; analysis of his application of understandings better measures his assimilation and accommodation of concepts. The test instrument designed to meet these criteria is the *TAB Science Test.* Using the tab-item format this test samples the behaviors of inquiry as denoted in the inquiry model.

Two parallel forms of the TAB Science Test were developed. The four sections of each test are:

Section 1

In Section 1 the child is presented with a science problem in the form of a physics problem-focus film. He is then asked to select from a set of explanations the one he thinks to be most correct. These explanations were selected from oral explanations given by children in classroom situations. In Section 1, the child either selects a correct answer or an incorrect answer. Knowledge of his selection gives some indication of his verification or discovery behaviors. If the child hypothesizes the correct solution to the problem before gathering clue data, he is verifying a solution. On the other hand, if the child does not know the correct solution to the problem before gathering clue data, any solution he arrives at will be a discovery.

Section 2

In Section 2 the child is presented with several (ten to sixteen) clue questions which he may use to help him solve the problem. The answer to each question is covered with a laminated tab attached to the test with double edged masking tape. If the child wishes to ask a question, he is free to remove the corresponding tab to find the answer. However, before he is allowed to remove any tabs, the child is instructed to read *all* of the clue questions. This precaution is taken to discourage the child from taking the "cafeteria" sequence; that is, taking tab A, B, C . . .

Each question in Section 2 was selected from a group of questions asked by children in similar problem situations. Previous to drafting the test, thirty one-half hour to one hour Inquiry Sessions with fourth-, fifth-, and sixth-grade children were held. These sessions were based upon the format developed by Suchman (1962). In the Inquiry Sessions children were shown the identical films found in the *TAB Science Test* and were asked to solve the problems after gathering clues by asking questions of the investigators. In answering the questions one basic restriction was imposed: the questions had to be so structured as to be answerable by a "YES" or "NO" response. This restriction eliminated open-ended questions and forced the children to focus and structure their queries. As Suchman (1962) indicated:

> For example, the child may not ask: "How did the heating affect the metal?" but he may ask: "Did the heating change the metal into a liquid?" In the first instance the child does not state specifically what information he wants. He is asking the teacher to conceptualize relationships for him, to teach him something. This is the very antithesis of what inquiry is designed to do. (Suchman, 1962, p. 30)

For this reason the questions in Section 2 of the *TAB Science Test* are answerable by "YES" or "NO" responses.

The clue questions on the *TAB Science Test* are those questions which were most often asked by children in the Inquiry Sessions and which were considered basic to the solutions of the problems. Each list of questions was analyzed by a panel of seven judges who tested each clue question for *scientific accuracy, question value* (Is the question relevant, additional, or irrelevant to the solution of the problem?) and *question type* (What kind of information is included in the question?)

In Section 2 analysis of the child's actions gives an indication of the child's searching and data processing behaviors. A child proficient at searching should not choose clues which are irrelevant to the problem. A child proficient at data processing should organize data into logical and/or non-redundant sequences of clue questions.

Section 3

In Section 3 the child is presented with the same series of explanations found in Section 1. However, in this section, to the right of each of the explanations is a numbered tab. When the child is ready to select one of the explanations, he is free to pull the tab and see if he is correct. A correct answer has a "YES" response under the numbered tab.

In Section 3 knowledge is obtained of the success or lack of success of the child's verification or discovery behaviors. If the child hypothesized the correct solution in Section 1 and then (immediately after gathering clues in Section 2) selected the correct solution again in Section 3, he is a more successful verifier than the child who selects an incorrect solution in Section 3 before deciding on the correct solution. This same trend holds for successful and non-successful discoverers.

Section 4

At the end of each test two multiple choice questions are asked of the child. These questions are included in the test to evaluate the child's ability to transfer discovered or verified concepts. This section, then, gives a measure of assimilation and accommodation behaviors.

A MEANS FOR EVALUATING PERFORMANCE ON THE TAB SCIENCE TEST

A crucial aspect of the development of the *TAB Science Test* was the identification of a scoring technique that reveals specific inquiry behaviors. In past studies there have been suggestions for scoring similar instruments that when put in language appropriate for describing the *TAB Science Test* are included in the following list:

(1) Comparing the order of tab pull of children with those who are proficient in inquiry or against an "optimal sequence" of tab pull.
(2) Use of the sum of "utility scores" of tabs (Rimoldi, 1955) which are defined as ratios

between the number of times that individual clue items are selected and the total number of subjects in the test sample.

(3) Scoring based on the number of tab questions chosen by the child before selecting the final correct answer, or use of an average or minimum number of tabs as a criterion.

(4) Using a weighted error count, where the clue tabs are weighted in inverse relation to their relevance in isolating the answer to the problem.

In this study the weighted error count that was proposed by Glaser, Damrin, and Gardner (1952) was selected as the major basis behind the scoring of the *TAB Science Test.* In this method, the child is penalized for choosing irrelevant and additional information clues more so than he is for choosing relevant clues.

Another important aspect considered in the scoring procedure of the *TAB Science Test* is the order of tab pull. While it is recognized that not all problems may be solved in a similar manner, it is important that logic guide the inquiry behaviors of children regardless of what method is used to solve a specific problem. For example, a child inquiring into part A-1 of Form A of the *TAB Science Test* finds the following questions in Section 2.

Question B: Was there water in the bottle with the yellow balloon before the bottle was put on the fire? YES

Question I: Did the bottle with the yellow balloon contain something besides air? YES

It is understandable that if the child asks Question I, a logical choice to follow would be Question B. However, if he asks Question B, then Question I would add nothing to his storehouse of knowledge and, in fact, would be illogical.

A computerized scoring system was developed to score the *TAB Science Test.* This scoring system takes into consideration the number, type and value of clues obtained by the child. Also taken into consideration is the order of tab pull. Flow sheets such as that in Figure 2 were developed for each test problem. On these

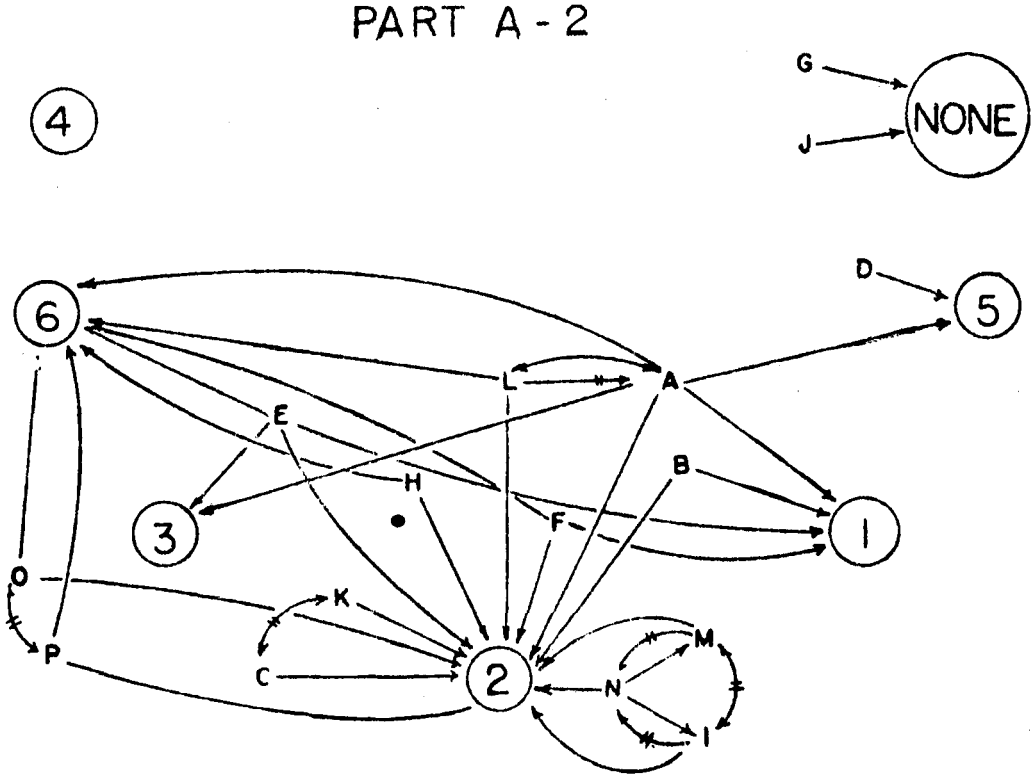

FIG. 2. System Analysis of TAB Science Test

sheets letters indicate the clue questions and numerals indicate the problem solutions. Arrows indicate possible connections between clue questions and problem solutions and between clue questions and other clue questions. Arrows which are blocked (—|–|→) indicate an illogical clue gathering sequence. Double pointed arrows (←|–|→) indicate redundancies.

Low numerical values are assigned to relevant clue questions, to logical and nonredundant sequences, and to plausible problem solutions. Subtraction of the sum of all tab and sequence scores counted against a child from a constant gives the final score. The constant is the maximum possible number of points that can be counted against any scoring sequence.

The following is a description of the analysis of each child's *TAB Science Test* behaviors. For each child tested, a page of data is available in a format exemplified by Figure 3. The student's name and an identification number for his class, sex, and *TAB Science Test* form are found at the top of Figure 3. Below this identification are two graphs. The graph at the top of the figure represents the first problem of the *TAB Science Test* Form B (The Air Sled). Bordering the graph as the abscissa are numerals 1 through 14. Bordering the graph as the ordinate are numerals 1 to 4 and letters A through J.

A point (asterisk) on this graph indicates the specific clue tab that was selected by the student. For example, on Figure 3, it can be noted that the first clue tab selected by this child was TAB E. After TAB E was selected, other tabs selected were TAB J, TAB A, TAB C, TAB G, TAB D, TAB H, and TAB B. After pulling these lettered tabs, the child thought that he knew the correct solution to the problem and pulled TAB 2 from the answer page. Under TAB 2 was a "YES" response, so the child knew he had found the solution to the problem.

The same format as described above is followed in interpreting the graph found at the lower half of Figure 3. This lower graph describes student behavior on *TAB Science Test* Form B—Part 2 (The Bimetallic Strip).

At the lower left of Figure 3 are six numbers:

*27**
68 (10, 0)
95
266

The *27* represents the number of points counted against the child by invoking the test scoring system as represented in the upper graph. This number is the sum of (1) the scores assigned to each tab [TAB E =1, TAB J=1, TAB A=1, TAB C=4, TAB G=3, TAB D=5, TAB H=1, TAB B=1, TAB 2=5] and (2) the scores assigned to illogical or redundant sequences chosen by the child [TAB E followed by TAB A is illogical; hence, E→A=5 points].

The 68 represents the number of points counted against the child by invoking the test scoring system as represented in the lower graph.

The (10, 0) indicates the child's proficiency at transferring concepts into other similar problem situations. There are two transfer questions on the *TAB Science Test* and, if the child answers a question correctly, he is assigned a score of 0. If he answers the question incorrectly, he is assigned a score of 10. In the case represented on Figure 3, the child missed the first question, but correctly answered the second.

The 95 is the sum of *27* and 68.

The *266* is the difference between 95 and a constant 361. Since the total number of points that can be counted against a person on Form B of the *TAB Science Test* is 361, the examinee's final score is obtained by substracting the sum of the top graph score (*27*) and the bottom graph score (68) from this constant.

The asterisk indicates that for the problem described by the top graph, the child hypothesized the correct solution before gathering any clues. In the second problem

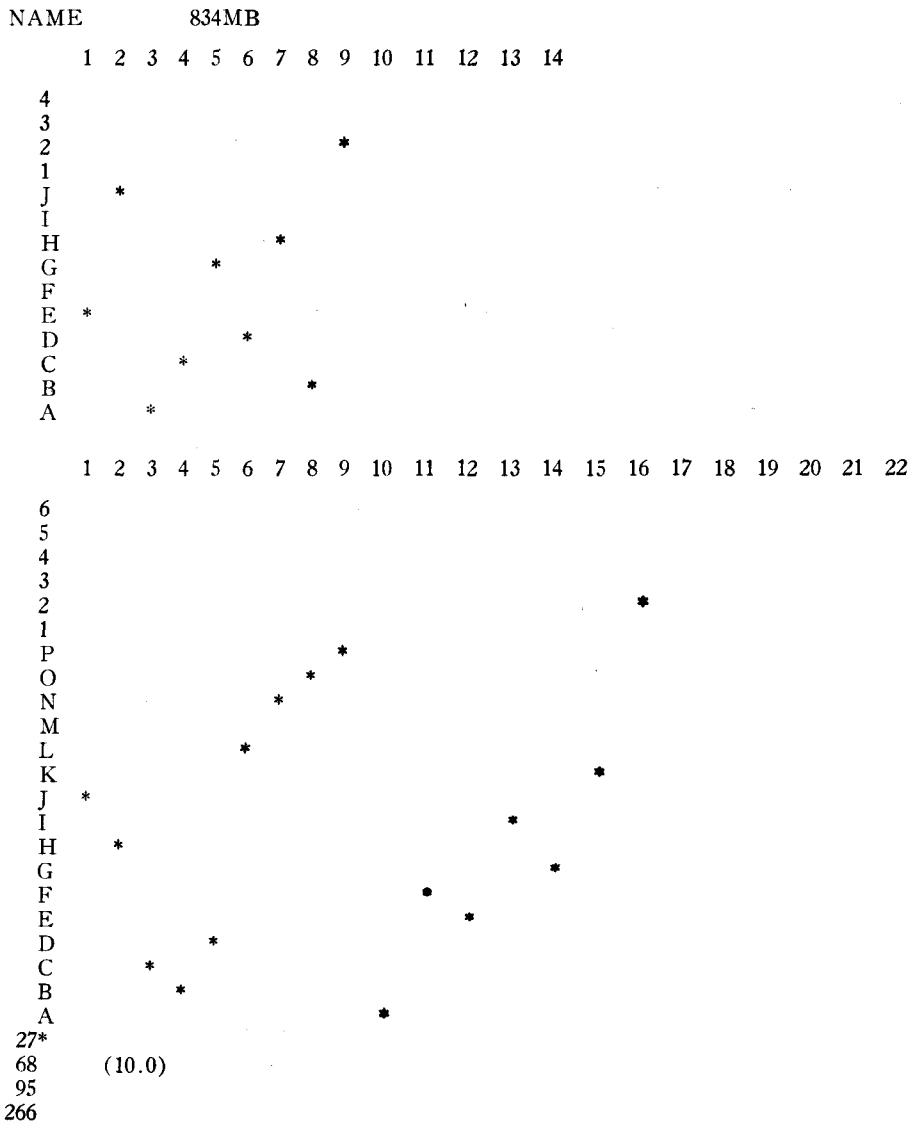

FIG. 3. The TAB Science Test Computer Print-Out.

(bottom graph), the child did not correctly hypothesize the problem solution, hence, no asterisk is included with his score. This asterisk gives one indication of the child's verification or discovery behaviors. If the asterisk is present, it is assumed that the child is verifying; if the asterisk is missing, then any correct solution that the child finds will be a discovery.

THE SAMPLE TESTED

The subjects tested with the *TAB Science Test* were 2519 fourth-, fifth-, and sixth-grade students in six Texas Independent School Districts. In each of these school districts a wide range of socio-economic backgrounds, tested intelligence, science knowledge, and reading scores were exhibited by the subjects. In all cases, the subjects were administered the *TAB Science Test* in groups of not less than 25. Care was taken to insure that all of the subjects were tested under standard conditions.

Data collected from *TAB Science Test* scores are presented in TABLE I.

BASIC TEST INFORMATION

The *readability* of the *TAB Science Test* was tested by using the Dale-Chall and the Spache Readability Formulae. Each form of the test was found to be acceptable for children having a fourth-grade reading vocabulary.

Validity

While construct validity can be invoked to describe the validity of an instrument such as the *TAB Science Test* an attempt was made to substantiate test norms by means of concurrent validity analysis. Concurrent validity of a test demands some external criterion. In the case of inquiry this difficulty is magnified—little has been done in the past with objective measurements of inquiry. A search of the literature revealed no test instrument which gives indication of the procedures of children as they solve problems.

In this study two external criteria were selected: (1) the science reasoning questions of the *Sequential Test of Educational Progress* (STEP Science) and (2) teachers' ratings. Using the STEP Science reasoning questions as a criterion, a validity coefficient (approx. .05) indicated that (1) above was inadequate.

Teachers were asked to rank their students in terms of the criterion "systematic problem solvers". Rankings were obtained from paired comparisons using multiple rank analysis as described by Gulliksen and Tucker (1959). Six of the fifteen teacher rankings of students were significantly correlated with the rankings obtained from *TAB Science Test* scores. The highest rank correlation was .64.

Reliability

Reliability of the *TAB Science Test* was determined by calculation of coefficients of equivalence and internal consistency. Comparing 446 students' scores on both forms of the *TAB Science Test*, coefficients of equivalence were obtained:

$$r_{ab} = .420 \quad (N=238)$$
$$r_{ba} = .365 \quad (N=208)$$

Coefficients of correlation were also calculated between scores obtained on the first problem of each test with subsequent scores on the second problem of each test. The following coefficients of internal consistency were obtained:

$$r_{\text{form A}} = .497$$
$$r_{\text{form B}} = .532$$

TABLE I

TAB SCIENCE TEST DATA

Test Form	Number Taking Test	Range of Raw Scores	Possible Maximum Score	Possible Minimum Score	Mean Score	Stand. Dev.	Median Score
A	1264	16–364	364	0	296	51.5	310
B	1255	15–345	346	0	260	58.5	274

Discriminatory Power

Chi square analysis indicated that the scoring system of the *TAB Science Test* does differentiate among nonproficient and proficient verifiers, data processors, searchers, discoverers, assimilators and accommodators. As indicated in Tables 2 through 6, higher scoring students were found to be more successful in these individual inquiry skills than lower scoring students.

SUMMARY

In this study, the *TAB Science Test* was developed to measure student inquiry. Chil-

TABLE II

COMPARISON OF SEARCHING BEHAVIORS WITH TAB SCIENCE TEST SCORES

	Students Choosing No. Irrelevant Clue Tabs		Students Choosing Irrelevant Clue Tabs		
	Form A	Form B	Form A	Form B	
Students with Highest 50 Scores	44	41	6	9	$x^6_{\text{Form A}}=51.86$
Students with Lowest 50 Scores	7	10	43	40	$x^2_{\text{Form B}}=36.01$
					$x^2_{.001}=10.83$

dren who were exposed to this test were exposed for the first time to an entirely different testing experience than that to which they were accustomed. What effect did test format have on the child's final score? Was the test measuring, in part, the degree reliability coefficients reflect in part the learning effect of the test format?

One measure of the concurrent validity of the *TAB Science Test* was the relationship between *TAB Science Test* scores and teachers' rankings of their students in terms

TABLE III

COMPARISON OF DATA PROCESSING BEHAVIORS WITH TAB SCIENCE TEST SCORES

	Students Choosing No. Redundant or Illogical Sequences		Students Choosing Redundant or Illogical Sequences		
	Form A	Form B	Form A	Form B	
Students with Highest 50 Scores	46	45	4	5	$x^2_{\text{Form A}}=60.94$
Stduents with Lowest 50 Scores	6	12	44	38	$x^2_{\text{Form B}}=41.78$
					$x^2_{.001}=10.83$

to which the child could follow directions and acclimate to a new test environment? Was the test administration a learning experience? Would *TAB Science Test* scores be more stable and consequently a more valid measure of inquiry if the child were instructed on the test format? Do the of the criterion, systematic problem solving. If teacher judgment of systematic problem solving—a criterion not usually of major interest in the assigning of grades to students—is such that the criterion is ambiguous and/or not completely agreed upon as a sing'e-dimension criterion, then further

TABLE IV

COMPARISON OF VERIFICATION BEHAVIORS WITH TAB SCIENCE TEST SCORES

	Students Demonstrating Successful Verification Patterns		Students Demonstrating Less Successful Verification Patterns		
	Form A	Form B	Form A	Form B	
Students with Highest 50 Scores	32	27	2	0	$x^2_{\text{Form A}}=29.66$
Students with Lowest 50 Scores	2	1	14	12	$x^2_{\text{Form B}}=35.17$
					$x^2_{.001}=10.38$

NOTE: The sum of the number of students showing discovery patterns plus the number of students showing verification patterns equals 100 for each form because there are two sub-problems in each test.

TABLE V

COMPARISON OF DISCOVERY BEHAVIORS WITH TAB SCIENCE TEST SCORES

	Students Demonstrating Successful Patterns		Students Demonstrating Less Successful Patterns		
	Form A	Form B	Form A	Form B	
Students with Highest 50 Scores	66	73	0	0	$x^2_{\text{Form A}}=117.96$
Students with Lowest 50 Scores	5	7	79	80	$x^2_{\text{Form B}}=143.12$
					$x^2_{.001} = 10.38$

study is needed to identify those factors that teachers consider as part of inquiry or problem-solving behaviors.

The results of test score analysis support the view that there are noticeable differences in individual student inquiry productivity as measured by the *TAB Science Test.* What student characteristics are definitive of productivity inquiry? What characterizes a student who scores highly on the *TAB Science Test?*

In the *TAB Science Test,* students are given the opportunity to gather clues in any order they choose. The scoring procedure of the test is based only upon the number of clues gathered and the logic of the order of clue gathering. Are there more productive patterns of clue gathering that evince the correct problem solution? Are these patterns related to student characteristics such as age, sex, reading ability, etc.?

Recent changes in some elementary school curricula have given added prestige to the role of student inquiry in the classroom. Since the *TAB Science Test* was designed to sample those behaviors involved in inquiry, it would be logical to assume that children who have been instructed in inquiry score more highly on the *TAB Science Test.* Further study is needed on the comparisons of student performance on the *TAB Science Test* with contrasting science curricula.

BIBLIOGRAPHY

Bloom, B. S., *et al. Taxonomy of Educational Objectives, Handbook I: Cognitive Domain.* New York: Longmans, Green, 1956.

TABLE VI

COMPARISON OF ASSIMILATION AND ACCOMMODATION BEHAVIORS WITH TAB SCIENCE TEST SCORES

	Both Transfer Questions Answered Correctly		First Question Correct Second Incorrect		First Question Incorrect Second Correct		Both Transfer Questions Answered Incorrectly	
	Form A	Form B	Form A	Form B	Form A	Form B	Form A	Form B
Students with Highest 50 Scores	28	25	11	10	11	15	0	0
Students with Lowest 50 Scores	1	3	12	13	11	6	26	28

$x^2_{\text{Form A}}=33.73$
$x^2_{\text{Form B}}=32.44$
$x^2_{.001} =16.27$

Butts, D. P. "The Relationship Between Problem-Solving Ability and Science Knowledge." *Science Education*, 49:138–146. 1965.

Damrin, D. E. "An Empirical Study of the Characteristics of the Problem-Solving Process." Paper delivered to American Psychological Association, 1953.

Glaser, R., Damrin, D. and Gardner, F. *The Tab-item Test—A Technique for the Measurement of Proficiency in Diagnostic Problem-solving Tasks.* Urbana: University of Illinois Press, 1952.

Gulliksen, H. and Tucker, L. R. *A General Procedure for Obtained Paired Comparisons from Multiple Rank Orders.* Princeton, New Jersey: Educational Testing Service, 1959.

Jones, H. L. "The Relationship Between Inquiry Training and the Problem-solving Behaviors of Elementary School Children." Unpublished Master's Thesis, University of Texas, 1964.

Piaget — — *The Psychology of Intelligence.* New York: Harcourt, Brace, and Company, 1950.

Rimoldi, H. J. A. "Technique for the Study of Problem Solving." *Educational Psychology Monogr.*, 15:450–461, 1955.

Sequential Tests of Educational Progress. *Teacher's Guide.* Princeton, New Jersey: Educational Testing Service, 1959.

Suchman, J. R. "Inquiry Training—Building Skills for Autonomous Discovery." *Merrill-Palmer Quarterly of Behavior and Development*, 7:147–169. 1961.

Suchman, J. R. *"Inquiry Training: Building Skills for Autonomous Discovery."* Urbana: University of Illinois, 1961.

Suchman, J. R. "Inquiry Training in the Elementary School." *Science Teacher.* 27:42–47, 1966.

Suchman, J. R. *"The Elementary School Inquiry Training Program."* Urbana: University of Illinois, 1962.

APPENDIX I

The films selected for the test were designed for the Illinois Inquiry Training Study for the explicit purpose of presenting problems to children. The chief advantages of the films are (1) they present data in graphic form and make use of continuous action to do so, and (2) they insure the uniformity of problem stimulus. In the two forms of the *TAB Science Test*, there are four films, one each for Parts A–1, B–1, A–2, and B–2. The films and the problems they pose are:

A–1. *Relative Expansion of Air and Water Under Heat.*
Two flasks are covered with balloons identical except for color. In one flask is air and a little water; in the other, there is only air. When the flasks are heated, the balloon covering the flask containing the water expands the greatest amount. The question is asked at the beginning of the film—"WHY DOES ONE BALLOON GET BIGGER THAN THE OTHER?"

B–1. *Air Sled.*
An air sled is glided over a level board. The sled moves more easily when the balloon is inflated and the rush of air from the balloon forms a cushion of air between the sled and the level board. When the balloon is deflated, there is considerably more resistance to movement. The question is asked at the beginning of the film—"WHY DOES THE BOARD MOVE EASILY AT FIRST BUT NOT LATER?"

A–2. *Boiling Water by Cooling.*
A pyrex flask is heated until the water inside the flask boils. The flask is then removed from the heat source and stoppered after the water ceases boiling. Cold water is poured over the outside of the stoppered bottle. The water inside the flask boils once again. The question is asked at the beginning of the film—"WHY DOES THE WATER BOIL THE SECOND TIME?"

B–2. *Bimetallic Strip.*
A bimetallic strip is heated and it bends. When emersed in cold water, it straightens. The question is asked at the beginning of the film—"WHY DOES THE BLADE BEND AND THEN STRAIGHTEN OUT?"

These films were selected for their clearness in presenting a problem, their relative difficulty, and the number of relational constructs necessary for their solution. Two films were selected for each form to insure a range of difficulty. Classroom experience indicated that the complementary films for each test form (A–1 vs. B–1; A–2 vs. B–2) represent nearly equivalent difficulty as problem stimuli for children in grades four, five, and six.