75-11,246

GRANT, Norris H., 1934-
THE VALIDITY OF OBJECTIVE TESTING AS A PROCESS OF APPRAISING THE THINKING ABILITY OF STUDENTS IN HIGH SCHOOL BIOLOGY AND PHYSICS.

The University of Oklahoma, Ph.D., 1974
Education, sciences

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

THE VALIDITY OF OBJECTIVE TESTING AS A PROCESS OF APPRAISING THE THINKING ABILITY OF STUDENTS IN HIGH SCHOOL BIOLOGY AND PHYSICS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

NORRIS H. GRANT

Bethany, Oklahoma

1974

THE VALIDITY OF OBJECTIVE TESTING
AS A PROCESS OF APPRAISING THE THINKING
ABILITY OF STUDENTS IN HIGH SCHOOL
BIOLOGY AND PHYSICS

APPROVED BY

[illegible]

[illegible]

[illegible]

Don Stafford

DISSERTATION COMMITTEE

THE VALIDITY OF OBJECTIVE TESTING AS A
PROCESS OF APPRAISING THE THINKING ABILITY
OF STUDENTS IN HIGH SCHOOL BIOLOGY AND PHYSICS

by: Norris H. Grant
Major Professor: John W. Renner

The primary objective of this study was to determine if a student's responses on multiple-choice items reflect understanding of formal and concrete concepts and if that understanding was commensurate with his level of mental development. By finding the relationship among the reasons given for selecting a particular response, the level of concepts used in the test items, and the thinking levels of the students, a determination of those concepts which could be understood by formal and concrete operational thinkers was possible.

Utilizing Piagetian-styled tasks, the operational levels of biology students were found to be 37% formal and 63% concrete and for the physics sample, 32% concrete and 68% formal. A positive correlation was found between formal and concrete-item test scores and reasons given by formal operational subjects in both samples. The correlation between concrete-item test scores and reasons given by concrete operational students was found to be positive for each sample. No correlation was found between formal-item test scores and reasons given by concrete operational biology and physics subjects.

Objective tests provide a valid measure of understanding formal and concrete concepts only when the operational level of each student is known. Students who can substantiate their responses on formal and concrete-concept items are operating at the formal level. These thinkers have the ability to develop an understanding of formal and concrete concepts. Objective test scores of concrete operational thinkers do not reflect understanding of formal concepts but do reflect understanding of concrete concepts. Concrete operational thinkers do not have the capacity to understand abstract or formal concepts but do have the mental ability to understand concepts developed through their senses or those concepts which can be experienced first hand.

ACKNOWLEDGEMENTS

To the members of the doctoral committee, Dr. Robert Petry, Dr. Don Stafford, and Dr. Robert F. Bibens, who gave of their time and valuable assistance, a sincere thank you is extended. Special gratitude is expressed to Dr. John W. Renner, Chairman of the doctoral committee, for his guidance through the graduate program and instruction during the preparation of this dissertation. It has been a plesant and rewarding experience working with Dr. Renner. His friendship is greatly cherished.

Gratitude is also extended to Marilyn Black, Linda Johnson, and John Knight for allowing their students to be tested at Putnam City High School.

For their time spent judging the test items on the subject matter examinations and helpful comments, sincere appreciation is given to Dr. Robert Karplus, Dr. Mary Budd Rowe, Dr. A.E. Lawson, Dr. A.E. Lawson, and Dr. Don Stafford.

Without the constant support and encouragement of my wife, Rosemary, completion of the graduate program and preparation of this dissertation would have been much more difficult. Her gracious endurance is acknowledged and appreciated.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Statement of the Problem

The orgin of this study may be found in the responses to the following letter printed in the Times of London:

> Sir,--Among the 'odd one out' type of questions which my son had to answer for a school entrance exam was: 'Which is the odd one out among cricket, football, billiards, and hockey?'
>
> I said billiards because it is the only one played indoors. A colleague says football because it is the only one in which the ball is not struck by an implement. A neighbor says cricket because in all other games the object is to put the ball into a net; and my son, with the confidence of nine summers, plumps for hockey 'because it is the only one that is a girl's game.' Could any of your readers put me out of my misery by stating what is the correct answer, and further enlighten me by explaining how questions of this sort prove anything, especially when the scholar has merely to underline the odd one out without giving any reason? [1]

The objective of this study was to determine if responses on objective test items accurately reflect student understanding of formal and concrete operational concepts and if that understanding was commensurate with each student's level of intellectual development as measured by utilizing Piagetian techniques.

To accomplish this objective, each of 89 high school biology and 59 physics students was given a multiple-choice examination in his respective subject area. That examination was composed of items which evaluated the examinee's understanding of concrete and formal concepts.

---

[1] Banesh Hoffman, The Tyranny of Testing, (New York: Collier Books, 1964). p. 17.

In addition to selecting a response to each question, each student was asked to give a written reason that would explain the response selected to each item. The student's level of mental development -- formal or concrete operational -- was determined by administering Piagetian-styled tasks.

This study was done to provide answers to the following questions. (1) Was there a relationship between the correct response students gave to a multiple-choice item and the reason(s) given for its choice? (2) What was the relationship among the reasons given for selecting a particular response, the level of multiple-choice items, and the intellectual level of the student, i.e., were concrete operational students able to give valid reasons only to concrete-concept items, whereas, could formal operational students give valid reasons for both concrete and formal-concept items?

Theory Base for the Study

In the past decade there has been a shift in the direction of science education in secondary schools.[2] To cope with the accelerated rate in which knowledge has accumulated, science educators have discouraged the memorization of facts and have encouraged student interaction with the environment. The acquisition of factual knowledge, hopefully, became a by-product of this action and not exclusively the result of content transmission by the teachers.

This new thrust of science education has clearly placed the burden of learning on the student by requiring him to analyze and interpret data, and make generalizations. Specifically students are expected to learn by interacting with a variety of situations and as a result of those interactions develop concepts and think logically with them. Furthermore, all students are expected to be able to participate in working toward that broad, general goal.

---

[2]Evidence for this assertion can be verified by the number of new course materials and curricular innovations published between 1959 and 1974 which emphasize the inquiry approach.

Have all students been able to perform as educators wished? Even with those curricula reforms that have provided students with an opportunity to interact with their environment, educators have not been totally satisfied with the results. As Kohlberg and Gilligan recently pointed out, "... some of the most enlightened proponents of the new curricula became discouraged as they saw only a sub-group of the high school population engaging with it."[3] This means that a portion of the students are not performing as educators desired and a reason may be that the thinking levels of the students whose performance was disappointing were quite different from that sub-group who could utilize the approaches found in the new curricula and develop an understanding of the concepts found there.

In a recent study Lawson[4] identified both concrete and formal operational thinkers in high school biology and physics and assessed the students' level of mental development by using Piagetian-styled tasks. The results showed that the thinking patterns of these students did influence their ability to develop understanding of certain concepts taught in these two subjects.

A possible explanation for Lawson's finding will be found in the work of Jean Piaget.[5] Piaget has shown that some students think in terms of reality while others can extend reality and think in terms of unlimited possibilities. As he has explained it:

> The adolescent differs from the child above all in that he thinks beyond the present. The adolescent is the individual who commits himself to possibilities... In other words the adolescent is the individual who begins to build 'systems' or 'theories' in the largest sense of the term.

---

[3]Lawrence Kohlberg and Carol Gilligan, "The Adolescent as a Philosopher: The Discovery of the Self in a Postconventional World," Daedlus, 1971, vol. 100, p. 1082.

[4]Anton Lawson, "An Investigation into Secondary School Science Curricular and the Intellectual Level of the Learner," Unpublished Ph.D. dissertation, The University of Oklahoma, (1973), p. 64.

[5]Bärbel Inhelder and Jean Piaget, The Growth of Logical Thinking from Childhood to Adolescence, (New York: Basic Books, Inc., 1958).

> The child does not build systems... The child has no power of reflection, i.e., no second-order thoughts which deal critically with his own thinking.[6]

This concrete thinker is unable to build theories because his thoughts are tied to reality, and he has to rely on first-hand experiences as a basis for developing concepts. He may, for example, understand the meaning of distance and time because his idea of these concepts are real to him. To take these ideas of distance and time and develop the concept of acceleration at this stage of his mental development requires mental operations[7] which he is unable to perform. To develop the concept of acceleration this thinker would not only need to think about an event-- the relationship between distance and time-- but also how this action is effected by periods of time between these events. He would have to superimpose one event on another by building a system of possibilities where reality is but one aspect. Since this individual thinks with concreteness and does not build systems, Piaget has called him concrete operational and has described the range of his logic to that which deals with classes, relations, and numbers.[8] His ability to perform mental operations is limited to the manipulation of objects and data that are real and not verbally expressed theories. Since this thinker is tied to reality and has no power to think about his thoughts, his conceptual understanding must come from interaction with real things. In short, this thinker can develop understanding of those concepts which are defined as concrete concepts.

Concrete concepts are classified as concepts whose meaning can be developed from first-hand experience with objects or events. These concepts may arise through intuition in which the entire meaning of the concept is given through the senses such as the color "blue". Concrete

---

[6] Ibid., p. 339-340.

[7] Mental operations are interiorized rather than exteriorized actions which can proceed in both directions and never occur in isolation. It is a form of mental action that modifies reality through a process of transforming real things into a generalized structure.

[8] Inhelder and Piaget, op. cit., p. 335.

concepts may also arise by postulation, however, if so part of the meaning of the concept must be sensed or immediately apprehended. Examples are common objects such as chairs, tables and other persons and events such as time, distance, and variation.

The adolescent who has the power of reflective thought and can think about the consequences of his thoughts Piaget has called formal operational. Formal operational thinking is basically an extension of concrete operational thinking because it is thinking which uses reality as only one among many possibilities. This type of thinking which usually occurs in adolescence after the age of twelve or fifteen[9] is best characterized by what Piaget calls propositional logic. Propositional logic is a form of logic that permits formulation of theories and provides a means of testing them without returning to concrete experimentation.

The formal operational thinker has the ability to reason with abstract ideas such as acceleration because from a structural standpoint he can perform mental operations that enable him to see the implications of a velocity change during an interval of time. His ability to extend his thinking from the present reality to the realm of the possible through propositional logic enables this thinker to develop and understand what are classified as formal concepts.

Formal concepts are concepts whose meaning is derived through position within a postulatory-deductive system. Meaning is given to these concepts not through senses but through assumption or through their logical relationships within the system. To fully comprehend the meaning of a formal concept one must be able to utilize formal operations or operate logically in a hypothetico-deductive manner, because a formal concept is one the meaning of which in whole or part

---

[9]According to Piaget, "Intellectual structures between birth and the period of 12-15 years grow slowly, but according to stages in development. The order of succession of these stages has been shown to be extremely regular and comparable to the stages of an embryogenesis. The speed of development, however, can vary from one individual to another." See Jean Piaget, "Intellectual Evolution from Adolescence to Adulthood," Human Development, 15: 1-12 (1972), No. 1.

is designated by postulates of the deductive theory in which it occurs. In other words the concept's meaning is developed by virtue of the properties or relations assigned to it by the postulate or set of postulates within which it is a member.

Some formal concepts can be imagined using concrete models. The solid particle model of atoms is an example of this type of concept. The light model which utilizes the pattern of waves on water is another example. Since concepts such as these can be modeled using concrete objects, events, or symbols, concrete operational individuals can develop the ability to verbalize about such things. These thinkers are, however, denied a full comprehension of the concepts until development of formal operations which allows construction of the postulatory-deductive system within which the concept derives its full meaning.

In evaluating learner comprehension of formal concepts which rely heavily on concrete models, it is imperative to avoid an evaluation based solely upon comprehension of the concrete models. This type of evaluation fails to test for understanding of a concept's implications.

Because of Piagetian theory, this investigator believes that how concrete and formal operational thinkers respond to objective test items should be directly related to their ability to develop understanding of concrete and formal concepts. According to Piagetian theory, the students' responses to test items that contain concrete or formal concepts should be commensurate with their levels of mental development. Concrete operational students, whose mental ability is limited to operating on real events and objects, should respond correctly only to concrete-concept test items. Formal operational thinkers possess the mental abilities to manipulate abstract and concrete data. They should, therefore, respond correctly to formal and concrete-concept test items.

If concrete operational thinkers chose the correct response to a formal concept item, the probability is that they guessed[10] because

---

[10]There is also the possibility that the thinker has memorized a verbal stereotype model of the concept and recognized some part of the model in the test item.

they lack the mental ability to develop ideas necessary for understanding the concept. This would imply that their response, although correct, does not give a valid measure of concept understanding. Piagetian theory, therefore, provides a basis for determining the validity of objective tests to measure understanding of formal and concrete concepts.

A correct response to a formal or concrete-concept test item by formal operational students should indicate understanding of these concepts because these thinkers have the mental ability to back up their answer with logic.

Need for the Study

Using Piagetian techniques Lawson[11] identified concrete and formal thinkers in high school biology, chemistry, and physics and classified formal and concrete concepts in those areas. By using especially constructed objective tests in those areas he found that the students' level of mental development influenced their ability to develop understanding of concrete and formal concepts.

Karplus and Karplus, 1970,[12] and Karplus and Peterson, 1970,[13] using an Islands Puzzle and a Ratio Problem assessed the reasoning ability of a large population of children and adults. These studies reported that subjects in grades ten through twelve used the same levels of reasoning as those concrete and formal operational thinkers identified by Lawson in his study, even though the reasoning abilities were assessed by evaluating written explanations of specific answers to the two problems.

Utilizing the written explanations to the Islands Puzzle and Ratio Problem Karplus, et al, found that the majority of the subjects in grades below eleven did not use formal logic in solving the two

---

[11] Lawson, op. cit., p. 64.

[12] R. Karplus and E. Karplus, "Intellectual Development Beyond Elementary School I: Deductive Logic", School Science and Mathematics, May, 1970, pp. 398-406.

[13] R. Karplus and R.W. Peterson, "Intellectual Development Beyond Elementary School II: Ratio, A Survey." School Science and Mathematics, Dec., 1970, pp. 813-820.

problems. These results would tend to support Lawson's claim that the students' level of mental development influenced their ability to develop understanding of concrete and formal concepts. The question remains as to whether or not subjects responding to an objective test are able to give valid written explanations for selecting a particular response to concrete and formal-concept test items. If a formal operational thinker does understand formal concepts as evident by his responding correctly to a formal-concept test item, then he should be able to give valid explanations for selecting the correct response. If on the other hand, a student selects a correct response to a formal-concept test item and cannot adequately explain his response, the probability was that he was concrete operational and did not really understand the concept. This would imply that these thinkers who cannot understand formal concepts probably guessed the response to the formal-concept test item.

Therefore, this study intends to show whether or not the scores received from an objective test reflect understanding of formal and concrete concepts as demonstrated by the ability of students to give valid written explanations for their responses. If the correlation is positive between test scores to formal-concept test items and correct reasons given by formal operational thinkers and zero between test scores to formal-concept test items and correct reasons given by concrete operational thinkers, then Lawson's findings have validity.

Furthermore, if Lawson's study is valid then scores from objective tests that are made up of mostly formal operational concept test items are measuring the students' level of mental development rather than concept understanding. Unless teachers know beforehand the operational level of their students, their interpretation of objective test scores will have little validity in terms of whether or not the scores reflect thought and understanding of the testing material. Therefore, this study will emphasize the importance of knowing the operational level of the students as well as the level of concepts to be measured before teachers attempt to appraise students' understanding of these concepts by using objective tests.

The Substantive Hypothesis

Whether or not a student selected the correct answers to concrete and formal test items on a multiple-choice examination based on his understanding or guessing can be determined by his ability to give logical, written reasons that explain his answer. The ability of students to give correct reasons for their choices to the test items depends upon their level of mental development and the level of the concepts of the test items. That is, students who are formal operational thinkers can give correct written reasons to both formal and concrete-concept test items, but concrete operational students can give correct reasons only to concrete-concept test items. Therefore, objective tests are not valid measure of understanding of formal and concrete concepts unless the teacher understands the level at which the learner is thinking. Having the students substantiate their responses with a valid written reason enables the teacher to understand where the students are operating.

Delimitations

This study was limited to high school students enrolled in biology and physics at one large high school in the Oklahoma City area. The duration of the study was one semester after the subjects were taught two designated topics[14] in their respective subject area.

Assumptions

The major assumptions for this study were:

1. Formal and concrete concepts can be identified from the subject content taught during the designated period and these concepts can be incorporated in a multiple-choice examination for each subject area.

2. The validation of the test items to measure formal and concrete understanding can be accomplished by having a panel of judges rate the test items.

---

[14]The topics are "Adaptation and Populations" for biology and "Measurements and Rectilinear Motion" for physics.

3. The subjects' level of intellectual development can be assessed by utilizing four Piagetian-styled tasks.

CHAPTER II

DESIGN AND PROCEDURE OF THE STUDY

Subjects

Subjects that comprised the biology and physics samples for this study were selected from one high school in the Oklahoma City area. This school has an enrollment of 3,000 students and three physics and 15 biology sections. Students are computer-assigned to the various sections of a course.

In the physics sections, students used Physics: A Basic Science[1] with laboratory activities taken from the laboratory manual developed by the Physical Science Study Committee.[2] Approximately one-half of the biology sections used Modern Biology,[3] and the other sections used Biological Science: An Inquiry into Life.[4] In both sections, the laboratory exercises were taken from the laboratory manuals accompanying the respective text-book used in the courses.

The approach used to teach the courses depended upon the individual instructors, but from personal interviews with each instructor, the investigator concluded that a combination of traditional lecture and some inquiry techniques were employed. The same instructor taught all three physics classes, but five different instructors taught the biology classes.

---

[1]Frank Verwiebe and Gordon Van Hooft, Physics: A Basic Science, (New York: American Book Company, 1970).

[2]Physical Science Study Committee, Physics, Second Edition, (Boston: D.C. Heath and Company, 1965).

[3]James H. Otto and Albert Towle, Modern Biology, (New York: Holt, Rinehart and Winston, Inc., 1969).

[4]Biological Science Curriculum Study, Biological Science: An Inquiry into Life, Yellow Version, (New York: Harcourt Brace, 1968).

The physics sample consisted of 59 subjects which comprised the entire population enrolled in the three sections. The biology sample consisted of 89 subjects enrolled in three different sections. The three sections used were selected at random[5] from the 15 different sections that were taught. Two of the biology sections selected used Modern Biology and the other used the Biological Science Curriculum Study materials.

MEASURING INSTRUMENTS

Two measuring instruments were used to gather data for assessing the subjects' level of mental development and providing test scores and reasons. They were (a) the Piagetian-styled tasks (b) the subject matter examinations.

Piagetian-styled Tasks

The following four tasks patterned after Piaget and Inhelder's work were given to each subject to determine his level of intellectual development. When data from these tasks are analyzed using principal components analysis, all of the tasks correlate highly with the same component.[6] Since the concrete operational tasks used by Lawson and Renner correlated highly with a different component, the conclusion was drawn that the first component was formal thought.[7]

The Piagetian tasks were administered to each subject and the responses solicited in individual interview conducted by the investigator. To insure that the results were consistent with previous research where these tasks were used, the investigator followed the same techniques for scoring and administering the tasks as those used by Renner and Lawson.[8] The scores assigned to the various levels of

---

[5] This was accomplished by using a table of random numbers and selecting the third digit from entries in the fourth line and second column until three different sections numbered from 1 to 15 were chosen. See, C.R.C. Standard Mathematical Tables, Chemical Rubber Publishing Company, (Cleveland, Ohio: Twelfth Edition, 1959), pp. 237-243.

[6] Anton E. Lawson and John W. Renner, "A Quantitative Analysis of Responses to Piagetian Tasks and Its Implications for Curriculum", Science Education, Vol. 58, No. 3. In Press.

[7] Ibid.

[8] Ibid.

performance on the tasks were as follows:

Level	Score	Description of Level
IIA	1	Beginning concrete operational
IIB	2	Concrete operational
IIIA	3	Beginning formal operational
IIIB	4	Formal operational

Description of the Piagetian Tasks

1. Conservation of Volume Using Two Identically Shaped Cylinders of Different Weights

In this task the subject was given two test tubes of water filled to the same level and two identical metal cylinders with obviously different weights. The subject was asked to predict how the levels of water in the two tubes would change after the metal cylinders were lowered in each. If the subject gave the correct response and could explain why, he was rated IIIA which was the maximum level of achievement for this task. The subject who predicted that one cylinder (usually the heavier one) would raise the level of water higher than the other one, but could explain the discrepancy after seeing that the levels were the same was rated IIB. The subject who was unable to explain his incorrect response was rated IIA.

2. Separation of Variables

The flexibility of a rod depends upon several properties including the material of which it was made, its length, its thickness, and its cross-sectional form.

The subject was given an apparatus consisting of a set of six rods differing in composition (steel and brass), length, thickness, and cross-sectional form, (round, square, and rectangular). Three different weights could be attached to the ends of the rods. The rods were attached to the edge of a platform in a horizontal position, in which case the weights exerted a force perpendicular to the platform. The subject was asked to identify the variables, then separate out experimentally the relevant variables that caused one rod to bend more than the others. To do this he had to place explicit multiplicative schemes together (thinner and longer), and to finally form a hypothetico-deductive reasoning scheme. With this scheme he actively attempted to verify and correctly determined which rod would be more or less flexible. One who performed the operation and gave correct reasoning for his prediction was rated at the IIIB level. The subject who was able to determine which rod was more or less flexible but failed to demonstrate how one variable could be held constant while allowing the others to vary in the multiplicative scheme in order to separate the effects of each variable was rated IIIA. A IIB rating was given to the subject who was unable to construct a hypothetico-deductive scheme as evident by his random technique of comparing the effects of each variable. The subject who could not perform any of the tasks was rated IIA.

3. Equilibrium in the Balance

The subject was shown an apparatus which consisted of a balance arm with 30 holes drilled in one inch intervals along its length. Weights of 10, 5 and 2 units,[8] which could be hung in the holes, were also pointed out. The examiner hung one 10 unit weight 7 inches from the fulcrum and asked the subject to hang another 10 unit weight on the opposite arm to achieve a balance.

Following completion of this, the subject was given two 5 unit weights to replace one of the 10 unit weights and asked to place them on the arm to achieve balance. While the subject determined the proper location for the 5 unit weights, the examiner held the arm of the balance parallel to the table surface. After the subject hung the weights, the examiner released the arm to see if a balance had been achieved.

Next the subject removed his weights, was given one 5 unit weight, and asked to hang it to again balance the 10 unit weight. Again the examiner held the end of the balance arm. After the subject had selected the location, he was asked to explain his selection. The subject who gave the correct position and explanation was rated IIIA. Following the explanation, the examiner let go of the arm to again see if balance

---

[8]The units could represent any combination of weights as long as the 10:5:2 ratios were maintained. For example, 200 grams was used for the 10 unit weight, 100 grams for the 5 unit and 40 grams for the 2 unit weights.

had been achieved. If the subject reasoned that the 5 unit weight was less than the 10 unit weight, therefore, it must be placed 5 inches farther out, he has used the difference between 10 and 5 for his comparison, not the ratio of 10 and 5. This response was rated at the IIB level. No systematic explanation or response which indicated the subject does not understand the law "heavier$\rightleftharpoons$near," placed the subject at the IIA level.

In a final task to differentiate IIIA from IIIB responses, the subject was given a 7 unit weight and asked to make it balance a 10 unit weight placed 7 inches from the center. Correct prediction and explanation of this proportion problem placed the subject at the IIIB level.

4. Operations of Exclusion

In this task which involves the oscillation of a pendulum, the subject was given several irrelevant variables along with the relevant variable. He had to discriminate between the appropriate variable by means of testing and to eliminate those variables which had no effect upon the rate of oscillation.

The subject who could not separate any of the variables was rated IIA. This subject was able to serially order the length of string, elevation of the mass and impetus but could not consistently order the mass of the pendulum bob. The subject who could separate only the push given the mass when starting out was rated IIB. This subject recognized that shortening and lengthening the string made the bob swing faster and slower but could not exclude the effect of mass and angle of release. The subject who was able to recognize the causal role of the length of the string and excluded the irrelevant variables by holding the length of string constant while allowing the others to vary was rated IIIA. A IIIB was awarded to the subject who recognized the causal role of the length of the string on the oscillation and could verbalize his actions for a general condition for all pendula and demonstrate how it could be tested.

Subject Matter Examinations

Multiple-choice examinations were written in the subject areas of biology and physics. The subject content used for writing the test items was chosen from units that were taught at the beginning of each course. The topics used for writing the biology examination were "Populations and Adaptation." The physics examination was written from material on "Measurements and Rectilinear Motion."

1. Concept Validation

A preliminary form of each examination was sent to a panel of judges to validate the level of concepts contained in each item as formal or concrete.[9] The criterion used for rating each item was the level of thinking ability needed to understand the meaning of formal and concrete concepts as they were defined in Chapter 1.

2. Experimental Tryout of the Items

Before the test items were assembled in the final form of the examination, a tryout of the test was given to samples of biology and physics students selected from students who had completed biology and physics the previous year. These students, who resembled as nearly as possible the subject who were to take the examination for this study, were given the tryout test during the 1973 spring and fall semesters at the high school where the samples for this study were selected. The physics tryout test was administered to all of his students from the previous year by the same instructor who gave the final form of the subject matter examination used in this study. The biology tryout test was given by this investigator to all students in his two chemistry and three electronics classes who had completed the course the previous year.

The purposes for the tryout were: (1) to identify defective or ambiguous items and non-functioning distractors; (2) to determine the difficulty and internal-consistency indices; (3) to provide information about the time needed to respond to the tryout items for determining the approximate number of items to be included in the final form of the examinations.

---

[9]See Appendix I and II for copies of the test items, the rating of each item by concept levels, and the qualifications of the panel members.

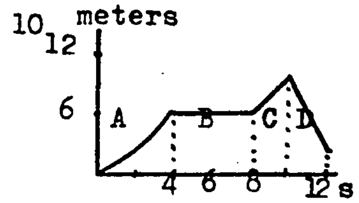
The samples who took the tryout were divided into high and low-scoring groups with each consisting of 27% of the sample.[10] The data accumulated from the tryout were assembled on index cards similar to the one illustrated below.

TABLE 1

INDEX CARD

SAMPLE OF CHOICE-BY-CHOICE RESPONSES TO PHYSICS ITEM 10

	Test: physics 1 Item 10 Sample A			Test: physics 2 Item 10 Sample B		
		high	low		high	low
	N	17	17		17	17
1	A	0	0	A	0	3
2	B	1	2	B	0	3
3	C	8	13	C	2	2
4	D	8	2	D	15	6
5	E			E		
6	omit	0	0	omit	0	0
7	NR	0	0	NR	0	0
8	% corr	29	0	% corr	80	25
9	D I	- 15		D I	55	
10	r	.57		r	.70	



Which portion of the graph indicates the greatest Av. Velocity?
a) A
b) B
c) C*
d) D

*Denotes correct answer

The information shown on the above index card was taken from the distribution of responses from two different tryout samples. Sample A was the third hour physics class and sample B the sixth hour class. The data under the high and low-scoring groups of sample A and B show the choice-by-choice distribution of responses made to the test item written on the left side of the card. For example, none of the high or low-scoring group in sample A marked choice "A" while three in the low-scoring group from sample B marked it. This kind of information was useful in detecting weak or non-functioning distractors. Also it can be seen that 15 in the high-scoring group from sample B marked the correct response while only 8 in this group from sample A marked it. This difference in number of correct

---

[10] This was done to facilitate the calculations and did not effect the statistics. See, E.F. Lindquist, Editor, Educational Measurements, (Washington, D.C.: American Council on Education, 1951), pp. 296-299.

responses was reflected in the difficulty index found on line nine which will be discussed in the next section. Besides the distribution of responses found in lines two through five, the number of examinees who omitted the item and who did not reach the item in the time allotted (NR) were recorded on lines six and seven. This information was valuable for determining whether the low-scoring groups were omitting the item more frequently than the high-scoring groups. If this were the case, the test item would be functioning as was expected. If on the other hand, more of the high-scoring groups were omitting the item, there would be indication that the item was ambiguous and needed rewording or should be deleted from the test. The frequency of NR and the position of the item in the test, in this case 10 for the above example, would indicate that the final form of the examination could contain at least this number of items.

3. Difficulty and Internal-consistency Indices

An analysis of the distribution of responses recorded in Table 1 also provided data for computing the indices of difficulty and intenal-consistency for each item. The index of difficulty which test writers simply define as the relative difficulty level of an item, theoretically should be around 50. For the example given, the index was found to be 15 for sample A and 55 for sample B. These values, found on line nine of the index card, are expressed in per cent of adjusted correct responses of examinees. The index of 55 was found by averaging the per cent of correct responses for the high-scoring group (80) and the low-scoring group (25). The index of sample A was found using the same procedure. The equation[11] for finding the per cent of correct responses is

$$P_t = 100 \frac{R_t - W_t/(k_i - 1)}{N_t - NR_t},$$

where P_t = the per cent of correct responses in the entire sample adjusted for chance success and for omission caused by not reaching the item in the time limit,

---

[11]*Ibid.*, p. 280

R_t = the number of examinees in the entire sample who answer the item correctly,

W_t = the number of examinees in the entire sample who answer the item incorrectly,

k_i = the number of choices in the item,

N_t = the number of examinees in the entire sample,

NR_t = the number of examinees in the entire sample who do not reach the item in the time limit.

The index of internal-consistency is the extent to which the item measures the same mental function as the total test. This index is recorded on line 10. The interpretation of the values for the above example meant that tryout sample A (.57) subjects gave responses to this item that were not as representative or consistent compared to the responses to the entire test as those given by sample B (.70). The index of sample B was higher because their responses had a higher correlation to the distribution of responses given on the entire test. The higher the index of internal-consistency, the more representative the item is in terms of measuring what the entire test was intended to measure, and the value of the index is analogous to a correlation coefficient. In other words, there is no accepted value for a good index because the magnitude of one value is not directly proportional to another value. A value of .50, for example, is not twice as good as an index of .25. Most authorities on test writing would consider that the item with the higher index, if two different items were written to measure the same mental function, would be the one that should be included in the test. Since both of the indices for the above example were for the same item and are considered acceptable, the information shows only differences in the distribution of responses between the two samples.

The equation used to calculate the index of internal-consistency was the bi-serial product-moment correlation.[12] That equation is

[12] *Ibid.*, p. 290

$$r_{pbis} = \frac{M_r - M}{\sigma}\sqrt{P_r/(1-P_r)}$$

where M_r = the mean criterion score[13] of examinees who marked the item correctly,

M = the mean criterion score of examinees who reached the item in the time limit,

σ = the standard deviation of criterion scores on examinees who finished the test,

P_r = the number of examinees who marked the item correctly divided by the number who finished the test.

4. Item Selection

Items that had the highest internal-consistency index and difficulty index of approximately 50 were used in the final form of the examination since research[14] has shown that the combination of these two factors provide the most valid measurement of the testing material and best discriminating ability among the majority of the examinees taking an objective examination. When an item has a difficulty index of 50 and at least a positive index of internal-consistency, theoretically the item will discriminate best between those who understand the item and those who do not. That is, if a test item with a difficulty index of 50 were given to 100 subjects, it would discriminate between 50% (50 subjects) who could answer and 50% (50 subjects) who could not answer the item making a total discrimination of 2500.[15] In terms of this research, a test item that was too easy (high index) meant that the majority of examinees would be able to answer the item and the reasons required would not extend their

---

[13]The criterion score is the raw score made on the test.

[14]Lindquist, op. cit., pp. 308-310.

[15]If the difficulty index were 40, the test would have 40x60 (2400) discriminations, whereas, a difficulty index of 70 would have 2100 discriminations. These calculations show that a test with most of the items at the 50 index level would have the maximum number of discriminations.

thought process to cover both formal and concrete operational levels of thinking.

On the other hand, an item with a low index (high difficulty factor) would not be answered by the majority of examinees and would discriminate against those who potentially understand the concepts but could not discriminate between the item distractors. In order for the scores on the objective test to have an equal chance to measure understanding of the different concepts as the reasons given, the level of difficulty of the items selected for the final form of the tests was chosen to be as near to 50 as possible.

5. Reliability

The central purpose of this study was to determine whether or not test scores from an objective test reflect the same understanding of concepts as the explanations the examinees gave for selecting the answers. In order to have faith in the ability of the multiple-choice examinations to reflect what the instruments were intended to measure (understanding of concrete and formal concepts), the tests' reliability was considered one of the more important aspects of writing the examinations. In addition, to be able to generalize the conclusions drawn from the results of this study to similar urban schools, the results should be relatively free of random or chance errors of measurement. These characteristics of reliability -- consistency of the results and accuracy of measurement -- were both considered when the finished form of the multiple-choice examinations was assembled and administered.

The reliability of the subject matter examinations was determined using the equation[16]

$$r_{tt} = \frac{n}{n-1} \left\{ \frac{S_t^2 - \sum_{i=1}^{n} p_i q_i}{S_t^2} \right\},$$

[16]Lindquist, op. cit., p. 587

where r_{tt} = the reliability of the total test,

n = the number of items in the test,

S_t^2 = the variance of the total test,

p_i = the proportion passing item i,

$q_i = 1 - p_i$.

6. Validity

Since the objective tests were constructed to measure understanding of formal and concrete concepts in biology and physics, the validity of the tests was equally as important as their reliability. The question of validity was whether or not the test was measuring what it was intended to measure. For this study, measuring understanding of different concepts was intended, therefore the emphasis was placed on validating the kinds of concepts contained in the test items. This validation, called construct validity, was accomplished by the panel of judges who rated the concepts in terms of constructs[17] assigned to the levels of ability needed to understand the different concepts. In this respect the validation was not quantified but was based on the judgement of the experts who were qualified to rate the levels of concepts.

7. Administering and Scoring the Examinations

The multiple-choice examinations were administered to each sample after a designated unit was taught by the instructors teaching the courses. The investigator reviewed the instructions for administering the tests which each instructor in order to assure that each subject would attempt to give written explanations for his responses. Each examination had specific instructions for the examinees pertaining to guessing and use of tables and slide rules.

The reasons given by the students for selecting a particular response to a multiple-choice item were judged valid or invalid by subjecting them to the following criteria.[18] Notice that the criteria

---

[17]These constructs were operationally defined as the ability to understand the meaning of concrete and formal concepts. See Chapter 1, pp. 4-6.

[18]See Chapter 3, p. 43 for examples of statements that were given by the students and how they were judged.

differ for the concrete and formal items.

1. Valid reasons for concrete concept items must contain:

a) Statements in which the subject clearly recognized data pertinent to the question asked and reported those data needed to explain the answer.

b) Statements in which the subject used his intuition to explain his answer. His intuition, however, was based on concrete data sensed from present reality or formed from first-hand experiences.

c) Statements in which the subject expressed postulations that led to correct explanations. These postulations were based on first-hand experiences and not hypothetical propositions.

2. Valid reasons for formal concept items must contain:

a) Statements in which the subject clearly showed evidence that he constructed a hypothetico-deductive scheme and used it to draw implications that correctly explained his answer.

b) Statements in which the subject inferred the use of a hypothetico-deductive scheme by giving the correct implications that explained his answer but did not specifically state the hypothetico-deductive scheme.

Statistical Treatment

The following hypotheses were tested for both samples of biology and physics subjects.

Ho_1: $r_{xy} = 0$. There is no correlation between the scores made on the multiple-choice test and the number of correct reasons given for formal and concrete concept items by formal operational subjects.

Ha_1: $r_{xy} > 0$. There is a positive correlation between the scores made on the multiple-choice test and the number of correct reasons given for formal and concrete concept items by formal operational subjects.

Ho_2: $r_{xy} = 0$. There is no correlation between the scores made on the multiple-choice test and the number of correct reasons given for formal concept items by concrete operational subjects.

Ha_2: $r_{xy} > 0$. There is a positive correlation between scores made on the multiple-choice test and the number of correct reasons for formal concept items by concrete operational subjects.

Ho_3: $r_{xy} = 0$. There is no correlation between the scores made on the multiple-choice test and the number of correct reasons given for concrete concept items by concrete operational subject.

Ha_3: $r_{xy} > 0$. There is a positive correlation between scores made on the multiple-choice test and the number of correct reasons given for concrete concept items by concrete operational subjects.

Notice that all of the alternate hypotheses were stated as positive correlations. The reason for giving the correlations as positive was due to the unlikely possibility that negative correlations could exist. If the correlations were negative, this would imply that there were valid explanations for incorrect selected responses.

Analysis of Data - The analysis of the data for testing the hypotheses involved two separate treatments.

1. The raw scores made on the multiple-choice tests were used to determine the level of confidence for deciding whether or not the results were significant. Rather than arbitrarily deciding between the .05 or .01 levels of confidence, the investigator allowed the scores made on the objective tests to set the level of confidence. If the .01 level had been chosen, the results of a small sample would have to be highly significant in order to accept the null hypotheses which could lead to committing a Type I error.[19] On the other hand, using a .05 level could lead to rejecting the null hypotheses because data were significant for a small sample and not for a large sample. This could lead to committing a Type II error.[20] Since the significance of data depends upon the size of the sample and the Chi-square test

[19]A Type I error is made when the null hypothesis is rejected when in fact it is true. The significance of each type of error will be discussed shortly.

[20]A Type II error is made when the null hypothesis is accepted when in fact it is false.

makes no assumptions regarding the shape of the distribution of traits measured by the criterion or item score, it was used to obtain the level of confidence. Using the results of this statistical treatment of the raw scores, the investigator was assured that the test scores from this study would reflect the same scores as those from a population large enough to be significant at the level of confidence found using the Chi-square test. The equation[21] used is

$$\text{Chi-square} = \frac{R_h - R_l - 1}{\sqrt{R_t\left(1 - \dfrac{R_t}{N_t - NR_t}\right)}},$$

where, R_h = the number of subjects in the high-scoring group who marked the items correctly,

R_l = the number of subjects in the low-scoring group who marked the items correctly,

$R_t = R_h + R_l$,

N_t = the number of subjects in the high and low-scoring group,

NR_t = the number of subjects in the high and low-scoring group who do not reach the item in the time limit.

Using the value from the Chi-square and Table 2, the level of confidence was found for the sample size used in this study.

TABLE 2[22]

SIGNIFICANCE OF LEVELS FOR CHI AT CERTAIN SAMPLE SIZES

Value of $R_t/2$	Values of Chi at Significance	
	.05	.01
2	1.90	----
6	1.93	2.46
7-10	1.94	2.52
37-100	1.96	2.57
101-over	1.97	2.58

[21] Lindquist, op. cit., p. 289.

[22] Ibid., p. 290.

2. The test scores and number of correct reasons were statistically treated using Pearson's product-moment correlation coefficient. Use of this statistic was possible because the numerical values given to the variables -- test scores and number of correct reasons -- permitted quantitative mathematical operations. Since Pearson's formula requires measurements of variables at the quantitative level and is one of the simplest to compute, it was used.

The computation equation[23] is

$$r_{xy} = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

where x and y are deviations from the means $\bar{X}$ and $\bar{Y}$ of the multiple-choice test scores and number of reasons given.

In order to determine the significance of the difference between the correlation coefficients for formal and concrete-concept test scores and reasons given by formal and concrete subjects, Fisher's Z_r transformation was used. The equation[24] is

$$Z_r = \frac{Z_{r1} - Z_{r2}}{\sqrt{\frac{1}{(N_1-3)} + \frac{1}{(N_2-3)}}},$$

where Z_r = the unit-normal curve deviate, with values of 1.96 and 2.58 required to be significant at the .05 level of confidence,

Z_{r1} = the Fisher transformation[25] of correlation coefficient r_1,

Z_{r2} = the Fisher transformation of correlation coefficient r_2,

N_1 = sample size of r_1, and N_2 = the sample size of r_2.

---

[23]George A. Ferguson, Statistical Analysis in Psychology and Education, (New York: McGraw-Hill, 1966), p.110.

[24]Ibid., p. 188.

[25]Transformations of correlation coefficients (r) to Z_r were obtained from a Table printed by Allen L. Edward, Statistical Methods for the Behavioral Sciences, Rinehart & Company, Inc., New York, see: Ferguson, op. cit., p. 412.

Level of Confidence and Types of Error

The null hypotheses were tested at the significant level determined by the Chi-square test. A summary of the possible error types are presented in Table 3 to show what the implications would be if incorrect decisions were made.

TABLE 3[26]

ERROR TYPES

(Ho = null hypothesis, Ha = alternate hypothesis)

		State of Nature	
		Ho is true	Ha is true
Decision	Reject Ho	Type I error	Correct decision
	Reject Ha	Correct decision	Type II error

Committing a Type I error, rejecting the null hypothesis when it was correct, would be the more serious error for the first and third hypotheses of this study. This would mean that the testing instruments were measuring what was expected but in reality there were no differences in ability of concrete and formal operational students to respond to different levels of concept items by giving valid reasons. This would imply that the whole concept of Piaget's theory was suspect. Committing a Type II error on the other hand, would be serious but only to the extent that the research was not designed to discriminate by testing the difference in abilities of these students to respond to formal and concrete concepts. The theory would not be in jeopardy, but the testing instruments either did not contain concrete and formal concepts, or the levels of the subjects were not correctly assessed.

How these error types would effect the outcome of this study is summarized in Table 4. These data provided the investigator with an overview of the entire study in terms of the significance of committing a Type I or Type II error.

---

[26] *Ibid.*, p. 163.

TABLE 4

SUMMARY OF COMMITTING TYPE I AND II ERRORS

Hypothesis	Error	Decision and Data Interpretation	Implications of Incorrect Decision
$Ho_1 : r_{xy}=0.$	I[a]	FOS[b] reasons for FCI[c] & CCI[d] and test scores are correlated: reject Ho_1	Data were as expected: theory was suspect or faulty
$Ha_1 : r_{xy}>0.$	II	Data not significant: accept Ho_1 - scores & reasons are not correlated	Test was not discriminating or was faulty: theory was as expected
$Ho_2 : r_{xy}=0.$	I	COS[e] reasons for FCI[c] and test scores are correlated: reject Ho_2	Data were not as expected: theory was as expected
$Ha_2 : r_{xy}>0.$	II[a]	Reasons and test scores are not correlated: accept Ho_2	Test was discriminating: theory was suspect
$Ho_3 : r_{xy}=0.$	I[a]	COS[e] reasons for CCI[d] and test scores are correlated: reject Ho_3	Data were as expected: theory was suspect or faulty
$Ha_3 : r_{xy}>0.$	II	Reasons and test scores are not correlated: accept Ho_3	Test was not discriminating or was faulty: theory was as expected

[a] More serious error in terms of this study

[b] Formal Operational Subjects

[c] Formal Concept Items

[d] Concrete Concept Items

[e] Concrete Operational Subjects

Committing a Type I or II error could lead to erroneous decisions about the hypotheses which could indirectly contradict Piaget's theory and invalidate this study. To minimize the errors, these measure were judiciously followed. (1) Concrete and formal concepts were accurately defined in terms of the thinking levels needed to develop and use each concept. (2) The levels of mental development were assessed carefully in terms of performance on the Piagetian tasks. (3) The written reasons given by the subjects had to demonstrate real understanding of the concepts and not merely an interpretation of what was expected by the investigator.

CHAPTER III

REPORTING AND ANALYZING THE DATA

The Subject Matter Examinations

Each form of the biology and physics examination contained 10 concrete and 10 formal multiple-choice items. The respective examination was administered to 89 biology subjects (Ss) and 59 physics subjects (Ss) during regular 55 minute periods. The specific instructions given the examinees were to respond to as many items as possible and give a written reason that would explain why a particular response was chosen. Only those items that received a response by all of the subjects are reported here and used in the analyses which follow. Table 5 gives the average number of correct responses ($\overline{X}$), average number of correct reasons ($\overline{Y}$), and the standard deviation (S.D.) for each subject matter examination. The average difficulty index (D.I.) for concrete and formal items, the reliability index (r_{tt}), and level of confidence (p) based on the Chi-square distribution are also reported in Table 5.

TABLE 5

SUBJECT MATTER EXAMINATION DATA

Examination	Ss	$\overline{X}$	$\overline{Y}$	S.D.	p	D.I. concrete	D.I. formal	r_{tt}
Biology	89	11.8	9.1	3.5	.01	67	41	.84[a]
Physics	59	11.7	10.8	3.3	.01	78	40	.87[a]

[a] $p < .01$

Piagetian Tasks

The level of mental development for each subject was assessed by analyzing his performance on four Piagetian tasks. Three tasks, equilibrium-in-the-balance, separation-of-variables, and operations-of-exclusion have four levels of achievement. These levels are pre-concrete (IIA), concrete (IIB), beginning formal (IIIA), and formal (IIIB). Students could score only at one of the first three levels on the conservation-of-volume task. If at least 75 per cent of a subject's reponses fell in either the concrete or formal category[1], he was placed there. Using this criterion and awarding numerical scores from one to four for successful completion of each level of achievement, a minimum score of 12 and a maximum score of 15 was possible for the formal operational level. If a subject was awarded ratings which totaled three IIIBs and one IIIA, he received a maximum score of 15 while one who rated four IIIA received 12. Both of these subjects were categorized as formal operational thinkers. Since the IIB level of achievement is considered concrete operational, subjects who received three IIBs and one IIIA (9 points) or four IIBs (8 points) were categorized as concrete operational. There were some subjects who received scores of 10 (two IIIAs and two IIBs) and 11 (three IIIAs and one IIB). These subjects who some investigators categorize as post-concrete[2] were assigned to the concrete operational level since this level was not considered different enough from concrete operational to be included as a special category. The combination of possible scores and the assigned level of each total score are summarized in Table 6.

---

[1]Jean Piaget, Judgement and Reasoning in the Child, (New Jersey, Littlefield, Adams and Company, 1966).

[2]John W. Renner, Don G. Stafford, and Ragan, William B., Teaching Science in the Elementary School. (New York: Harper and Row Publishers, 1973), p. 345.

TABLE 6

CLASSIFICATION OF ACHIEVEMENTS ON PIAGETIAN TASKS

Operational Level	Combination of Scores	Score	Level Assigned
Concrete	3 (IIB) +1 (IIIA) or less	≤9	Concrete
Post-Concrete	2 (IIIA)+2 (IIB) 3 (IIIA)+1 (IIB)	10 11	Concrete Concrete
Beginning Formal	4 (IIIA) 3 (IIIA)+2 (IIIB)	12 13	Formal Formal
Formal	2 (IIIA)+1 (IIIB) 3 (IIIB)+1 (IIIA)	14 15	Formal Formal

Those subjects in biology who scored 9 points or less comprised 37% of the sample compared to 17% for the physics sample. At the formal end of the spectrum, 42% of the physics sample achieved this level compared to 23% for the biology sample. The distribution of the post-concrete and beginning formal levels exhibited a similar pattern for each group. The frequency and per cents of subjects in each achievement level are reported in Table 7.

TABLE 7

LEVELS OF ACHIEVEMENT FOR THE PIAGETIAN TASKS

Sample	Concrete	Post-concrete	Beginning Formal	Formal
Biology	33 (37)[a]	23 (26)	12 (13)	21 (24)
Physics	10 (17)	9 (15)	15 (26)	25 (42)

[a] Numbers in parenthesis are per cents of the sample

The distribution of per cents of concrete and formal operational levels found in this study compared favorably with those reported by Lawson[3] whose study was done with students in the same grade and content area. That information is reported in Table 8.

---

[3]Lawson, op. cit., p. 64.

TABLE 8

COMPARISON OF TOTAL PER CENTS WITH LAWSON'S STUDY

Sample	This Study		Lawson's Study	
	Concrete	Formal	Concrete	Formal
Biology	63	37	64	36
Physics	32	68	34	66

Individual Scores, Reasons, and Levels of Mental Development

The individual's correct responses to the test items, the number of correct reasons given and his level of mental development are recorded in Tables 9 and 10. Those data are divided into correct scores for concrete (X_c) and formal (X_f) items and the corresponding number of correct reasons (Y_c) and (Y_f) for concrete and formal test items respectively. The subjects are ranked according to their total number scores on the entire test.

TABLE 9

INDIVIDUAL RAW DATA FOR BIOLOGY SAMPLE

Subject	T_i	X_c	Y_c	X_f	Y_f	Piagetian score	Level
1	20	10	10	10	10	14	F
2	19	10	8	9	8	15	F
3	18	10	10	8	8	15	F
4	17	10	10	7	8	14	F
5	17	10	10	7	7	14	F
6	17	9	9	8	8	14	F
7	17	10	10	7	7	14	F
8	16	10	8	6	6	14	F
9	16	10	9	6	5	14	F
10	16	8	9	8	6	14	F
11	16	9	9	7	7	14	F
12	16	9	8	7	4	13	BF
13	15	9	9	6	6	14	F
14	14	8	8	6	6	15	F
15	14	9	9	5	5	14	F
16	14	9	8	5	5	13	BF
17	14	8	8	6	6	13	BF
18	14	9	9	5	4	13	BF
19	14	7	6	7	3	11	PC
20	14	6	6	8	3	11	PC
21	13	7	8	6	6	14	F
22	13	7	7	6	6	13	BF
23	13	7	8	6	6	13	BF
24	13	9	9	4	4	13	BF
25	13	8	9	5	5	14	F
26	13	8	8	5	4	14	F
27	13	8	7	6	2	12	BF
28	13	8	4	5	2	11	PC
29	13	8	7	5	2	10	PC
30	13	9	9	4	2	10	PC
31	13	9	9	4	2	10	PC
32	13	7	6	6	4	10	PC
33	12	7	7	5	5	14	F
34	12	6	6	6	6	14	F
35	12	6	6	6	5	14	F
36	12	6	7	6	5	13	BF
37	12	5	5	7	7	13	BF
38	12	7	7	5	2	11	PC
39	12	5	4	7	3	11	PC
40	12	5	4	7	3	11	PC
41	12	6	6	6	1	11	PC
42	12	8	6	4	2	10	PC
43	12	6	5	6	2	9	C
44	11	6	6	5	2	12	BF
45	11	4	5	7	6	14	F
46	11	9	5	2	2	11	PC

TABLE 9 (continued)

Subject	T_i	X_c	Y_c	X_f	Y_f	Piagetian score	Level
47	11	9	9	2	1	10	PC
48	11	9	7	2	1	9	C
49	11	7	7	4	0	9	C
50	11	7	7	4	2	8	C
51	11	5	2	6	0	8	C
52	11	7	5	4	2	7	C
53	10	7	7	3	3	12	BF
54	10	7	6	3	1	9	C
55	10	8	8	2	1	9	C
56	10	8	7	2	1	9	C
57	10	6	6	4	4	11	PC
58	10	5	3	5	2	10	PC
59	10	6	5	4	0	8	C
60	10	3	2	7	2	10	PC
61	10	7	7	3	3	9	C
62	10	5	4	5	2	11	PC
63	10	5	5	5	1	8	C
64	10	7	7	3	0	9	C
65	9	5	3	4	2	11	PC
66	9	5	5	4	1	11	PC
67	9	8	6	1	2	10	PC
68	9	7	4	2	0	10	PC
69	9	8	6	1	2	10	PC
70	9	6	2	3	1	9	C
71	9	4	4	5	0	8	C
72	8	5	3	3	0	9	C
73	8	4	3	4	0	9	C
74	8	7	4	1	1	9	C
75	8	5	5	3	2	7	C
76	7	4	4	3	2	9	C
77	7	5	5	2	2	9	C
78	7	5	4	2	2	9	C
79	7	5	3	2	0	9	C
80	7	5	4	2	0	9	C
81	7	3	3	4	0	8	C
82	7	5	5	2	0	8	C
83	7	5	5	2	2	10	PC
84	6	4	3	2	0	7	C
85	6	4	4	2	0	8	C
86	6	3	3	3	2	10	PC
87	4	2	2	2	0	8	C
88	4	2	2	2	0	8	C
89	3	1	1	2	0	7	C

TABLE 10

INDIVIDUAL RAW DATA FOR PHYSICS SAMPLE

Subject	T_i	X_c	Y_c	X_f	Y_f	Piagetian score	Level
1	19	10	9	9	9	15	F
2	18	10	10	8	8	15	F
3	18	10	10	8	8	15	F
4	17	10	10	7	7	15	F
5	17	10	10	7	6	14	F
6	16	8	8	8	8	15	F
7	16	10	10	6	7	15	F
8	16	9	9	7	6	15	F
9	16	8	8	7	8	15	F
10	15	8	8	7	7	15	F
11	15	6	6	9	7	15	F
12	15	9	9	6	6	15	F
13	14	9	9	5	5	13	BF
14	14	9	9	5	1	14	F
15	14	8	8	6	7	15	F
16	14	10	10	4	4	13	BF
17	14	10	10	4	2	13	BF
18	13	9	9	4	2	14	F
19	13	10	10	3	3	13	BF
20	13	7	7	6	7	14	F
21	13	8	8	8	5	15	F
22	13	8	8	5	2	10	PC
23	12	8	8	4	4	15	F
24	12	8	8	4	3	14	F
25	12	9	9	3	2	13	BF
26	12	9	8	3	3	13	BF
27	12	5	4	7	7	13	BF
28	12	10	10	2	2	13	BF
29	12	8	7	4	2	12	BF
30	12	7	7	5	3	13	BF
31	12	9	9	3	2	13	BF
32	12	8	8	4	4	13	BF
33	12	8	8	4	3	12	BF
34	12	7	6	5	3	12	BF
35	12	8	3	4	1	10	PC
36	11	5	5	6	6	14	F
37	11	5	6	6	7	15	F
38	11	6	6	5	5	13	BF
39	11	7	7	4	3	12	BF
40	11	9	10	2	1	12	BF
41	11	8	7	3	2	11	PC
42	10	9	9	1	0	11	PC
43	10	9	9	1	1	10	PC
44	10	5	4	5	1	9	C
45	9	5	6	4	4	13	BF
46	9	6	6	3	3	13	BF

TABLE 10 (continued)

Subject	T_i	X_c	Y_c	X_f	Y_f	Piagetian score	Level
47	9	5	5	4	1	10	PC
48	9	8	6	1	0	8	C
49	8	5	5	3	2	11	PC
50	8	6	6	2	1	10	PC
51	8	5	5	3	2	10	PC
52	8	7	7	1	0	9	C
53	7	4	3	3	1	9	C
54	7	6	6	1	1	10	PC
55	7	6	4	1	1	9	C
56	7	6	5	1	1	9	C
57	7	4	3	3	0	9	C
58	7	5	2	2	0	9	C
59	6	3	3	3	1	9	C

Statistical Treatment

Correlation of Test Scores and Reasons- The product-moment correlation coefficient was used to calculate the statistics for the data presented in Tables 9 and 10. Correlations were found between the raw scores to each test item level and the reasons given by formal and concrete operational subjects. The mean scores and reasons for the concrete (A) and formal (B) portions of the objective tests, the results of the statistical treatment and levels of significance are reported in Tables 11 and 12.

TABLE 11

CORRELATION OF TEST SCORES AND REASONS FOR BIOLOGY SAMPLE

A. Results for Concrete items ($\bar{X}_c$) and Reasons ($\bar{Y}_c$)

Level	Ss	$\bar{X}_c$	$\bar{Y}_c$	Statistics	Significance
Concrete	56	5.9	4.9	$r_{xy} = .82$	$p < .005$
Formal	33	8.3	8.1	$r_{xy} = .88$	$p < .005$

TABLE 11 (continued)

B. Results for Formal Items ($\overline{X}_f$) and Reasons ($\overline{Y}_f$)

Level	Ss	$\overline{X}_f$	$\overline{Y}_f$	Statistics	Significance
Concrete	56	3.7	.89	$r_{xy} = .28$	$p < .02$
Formal	33	6.2	5.4	$r_{xy} = .81$	$p < .005$

TABLE 12

CORRELATION OF TEST SCORES AND REASONS FOR PHYSICS SAMPLE

A. Results for concrete items ($\overline{X}_c$) and Reasons ($\overline{Y}_c$)

Level	Ss	$\overline{X}_c$	$\overline{Y}_c$	Statistics	Significance
Concrete	19	6.2	5.3	$r_{xy} = .79$	$p < .005$
Formal	40	8.2	8.4	$r_{xy} = .94$	$p < .0005$

B. Results for Formal items ($\overline{X}_f$) and Reasons ($\overline{Y}_f$)

Level	Ss	$\overline{X}_f$	$\overline{Y}_f$	Statistics	Significance
Concrete	19	2.5	.89	$r_{xy} = .37$	$p = .05$
Formal	40	5.2	4.8	$r_{xy} = .85$	$p < .005$

<u>Fisher's Transformation</u> - Fisher's Z transformation was used to test whether or not the correlation coefficients between the biology and physics samples were significantly different. The procedure was to convert each correlation coefficient to Fisher's Z[4] and calculate the unit-normal-curve deviate Z. Values of Z of 1.96 are required for significance at the .01 level of confidence. Subscript 1 refers to the biology sample and subscript 2 to the physics sample. Each correlation coefficient had N-3 degrees of freedom.[5] The results are reported in Table 13.

---

[4]See Table E, Ferguson, <u>op. cit.</u>, p. 412

[5]The number of degrees of freedom is always the number of values that are free to vary, given the number of restrictions imposed upon the data. See Ferguson, <u>op. cit.</u>, pp. 156-157.

TABLE 13

TEST OF SIGNIFICANCE BETWEEN THE CORRELATION COEFFICIENTS OF THE BIOLOGY AND PHYSICS SAMPLES USING FISHER'S Z TRANSFORMATION

Correlations compared	r_1	r_2	N_1-3	N_2-3	Z_{r1}	Z_{r2}	Z
1) CI&R[a] by COS[b]	.82	.79	53	16	1.2	1.1	.3[e]
2) CI&R[a] by FOS[c]	.88	.94	30	37	1.4	1.8	1.3[e]
3) FI&R[d] by FOS[c]	.81	.85	30	37	1.1	1.3	.5[e]
4) FI&R[d] by COS[b]	.28	.37	53	16	.3	.4	.4[e]
5) FI&R[d] by COS[b] & FOS[c]	.28	.94	53	37	.3	1.4	4.2[f]
6) FI&R[d] by FOS[c] & COS[b]	.81	.37	30	16	1.1	.4	2.6[f]

[a] Concrete-concept item scores and reasons given.

[b] Concrete Operational Subjects.

[c] Formal Operational Subjects

[d] Formal-concept item scores and reasons given.

[e] $p < .01$

[f] $p > .05$

Interpretation of Statistical Treatment

Three hypotheses were formulated for this study. The first null hypothesis stated that no correlation exist between the objective test scores and number of correct reasons given by formal operational subjects to formal and concrete test items. The results for the biology sample showed that the correlation coefficients between the test scores and reasons were .88 and .81 for concrete and formal test items, see Table 11. For the physics group, the correlation coefficients were .94 and .85 for the same two variables, see Table 12. These statistical values indicated that the null hypothesis should be rejected at the .01 level of confidence and the alternate hypothesis should be accepted. That is, the correlation coefficients between objective test scores to concrete and formal concept items and the reasons given by formal operational biology and physics subjects are positive and high.

The second null hypothesis stated that no correlation exist between formal concept items and the number of correct reasons given by concrete operational subjects. The results that were found show correlation coefficients of .28 (Table 11 B) and .37 (Table 12 B) between tests scores and number of correct reasons given for formal concept items by concrete operational biology and physics subjects. These values were significant between the .05 and .02 level of confidence, therefore, the correlations were not significantly different from zero at the .01 level of confidence. These tests show the second null hypothesis should not be rejected.

The third null hypothesis stated that no correlation exists between the test scores and number of correct reasons given for concrete items by concrete operational subjects. The statistics revealed correlation coefficients of .82 and .79 for concrete operational biology and physics subjects respectively, see Tables 11 A and 12 A. These values were significant at the .005 level of confidence, therefore, the null hypothesis should be rejected. That is, the correlation between test scores and reasons given for this level of test item by concrete operational subjects was highly positive.

Test of Significance Between Correlation Coefficients - The results reported in Table 13 reveal the following information about those correlation coefficients that were compared. (1) There was no significant difference between those positive correlations that were reported in Tables 11 and 12 for the biology and physics samples. The information found on the lines 1-2 of Table 13 shows that the correlation coefficients between concrete test scores and reasons given by concrete and formal operational subjects for the biology sample (.82 and .88) and physics sample (.79 and .94) were not significantly different at the .01 level of confidence. Also on line three of the same table the positive correlations between formal test scores and reasons given by formal operational subjects from each sample were not found to be significantly different at the same level of confidence. (2) There was no significant difference between the non-positive[6] correlations of the biology (.28) and physics (.37) samples. The correlation coefficients that were compared were between formal test scores and reasons given by concrete operational subjects from each sample. (3) The difference between the positive (.94 for the physics sample and .81 for the biology sample) and the non-positive (.28 for the biology and .37 for the physics sample) correlations were found to be highly significant. This information found on lines five and six of Table 13 was for correlations between formal test scores and reasons given by formal and concrete operational subjects in each sample.

The results from the test of significance between the correlation coefficients of the biology and physics samples assured the investigator that the correlations used to test the null hypotheses represented the same measure of predictability for each sample. In other words, those positive and non-positive correlations represented the same magnitude of values in each sample. This would indicate that the positive correlations are measuring the same functional relationship between the variables for each sample. These variables for positive correlations are concrete and formal test scores and reasons given by concrete and formal operational thinkers respectively. Also, those non-positive correlations can be considered to represent the same nonfunctional relationship between the test scores and reasons given by each sample. In addition, those correlations that were not

---

[6]Non-positive correlations referred to here and in the following discussion mean those correlation coefficients that were not significantly different from zero and are not negative correlations.

significantly different for each sample could be used to predict performance of each operational level of thinker from either biology or physics on formal or concrete test items.

The difference between non-positive and positive correlations of each sample was found to be highly significant. This would support the claim that the .28 correlation coefficient for the biology sample and .37 correlation coefficient for the physics sample were not measuring a functional relationship between formal test scores and reasons given by concrete operational subjects in each sample. The positive correlations (.94 for the physics and .81 for the biology samples) being significantly different from the non-positive correlations would imply that these correlations were measuring a functional relationship between the formal test scores and and reasons given by formal operational subjects in each sample.

In sum, these findings supported the fact that (1) a positive correlation does exist between formal and concrete test scores and reasons given by formal operational biology and physics subjects; (2) a positive correlation does exist between concrete test scores and reasons given by concrete operational subjects in each sample; and (3) no correlation exists between formal test scores and reasons given by concrete operational biology and physics subjects.

Evaluation of Explanations

Criteria- The reasons given to explain each subject's selected choice to the multiple-choice items were evaluated by the investigator as valid explanations if they met the following criteria.

1. A correct explanation to a concrete concept item must contain the following kinds of statements.[7]

a) Statements in which the subject clearly recognized data pertinent to the question asked and reported those data needed to explain the answer.

---

[7]Examples of statements that met or did not meet these requirements are given in the section that follows.

b) Statements in which the subject used his intuition to explain his answer. His intuition, however, was based on concrete data sensed from present reality or formed from first-hand experience.

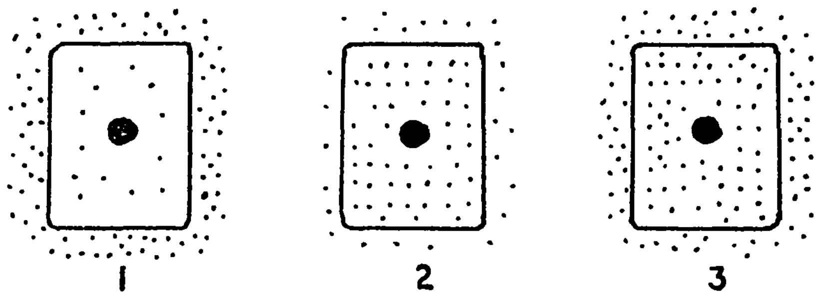
c) Statements in which the subject expressed postulations that led to correct explanations. These postulations were based on first-hand experiences and not from hypothetical propositions.

2. A correct explanation to a formal concept item must contain the following kinds of statements.

a) Statements in which the subject clearly showed evidence that he constructed a hypothetico-deductive system to draw implications that correctly explained his answer.

b) Statements in which the subject inferred the use of a hypothetico-deductive system by giving the correct implications that explained his answer but did not specifically state what hypothetico-deductive system was used.

Examples of Statements - The following reasons contain statements that were typical of correct and incorrect explanations and demonstrate how the investigator evaluated those reasons according to the foregoing criteria. The first example are reasons given for responses on a formal biology test item.

> The next question is based on the following information and diagrams of 3 cells in a water-sugar solution. The dots represent sugar molecules which cannot pass through the cell membrane.
>
> 
>
>
> "Which cell(s) will shrink in size?"

To answer the question required the construction of a hypothetico-deductive system in which the relation between concentration and diffusion could be used to predict which cell would shrink. A concrete model of concentration could be constructed, provided the meaning of concentration was understood, by counting the relative number of dots contained in each cell. Using the concrete model to identify which cell has the least concentration of sugar inside and the relationship between concentration and diffusion, a correct prediction was possible as to which cell would shrink. The ability to follow the line of reasoning expressed here was expected of formal operational thinkers, therefore, this item was rated formal operational.

The following reasons[8] were judged correct and reflected the thinking pattern needed to make these explanations.

1. "Only cell one has a smaller concentration inside, water will move to a higher concentration, therefore cell one will shrink."
2. "More water in cell one than out, water moves from lower concentration, therefore cell one will shrink."
3. "Water moves from a higher concentration [more water = higher concentration] to a lower concentration, this is cell one, it will shrink."

These reasons provided evidence that the subjects constructed a postulatory-deductive system and used it to explain why cell one would shrink. In these statements the subjects first identified the relative concentration of sugar in the cells -- an example where concrete data were recognized from the diagrams -- to establish a condition in which diffusion could operate. Secondly, these reasons show that the subjects expressed the relationship between concentration and diffusion, "water moves from a lower concentration," and concluded in a rational way that, "cell one will shrink."

It should be noted that understanding the relationship between concentration and diffusion was the basic ingredient for establishing the postulatory-deductive system. This relationship was more difficult to imagine because the concept of diffusion is generally taught as a process in which particules move from a higher to a lower concentration. In this particular problem the sugar molecules were not permitted to move through the cell membranes, therefore, a reverse relationship had to be imagined to solve the problem. The above reasons demonstrate that the subjects applied the

---

[8]The statements that each student gave are presented as he wrote them. Some of the statements contained incorrect grammar and spelling. Only the spelling was corrected by the investigator.

reverse relationship and arrived at the correct conclusion.

The following reasons were also judged correct and are examples of those kinds of explanations that fit the 2(b) requirements.

1. "Only cell one has smaller concentration than the solution, cell one will shrink."
2. "More water will move out of the cell to mix with the sugar outside of the cell in one."
3. "As the water diffuses out of the cell, its volume and size will shrink."

The primary difference between these reasons and those given in the first example is that no statements were found concerning the relationship between diffusion and concentration which would indicate that a postulatory-deductive system was constructed. The fact that these reasons hint at an implication, "... cell one will shrink," however, indicates that the thinkers may have recognized the results of diffusion but did not express the conditions in which this event would take place. These explanations were judged correct because the subjects' inferences were probably based on propositional logic even though the scheme was not verbalized.

The following example of reasons given to the same test item were judged incorrect.

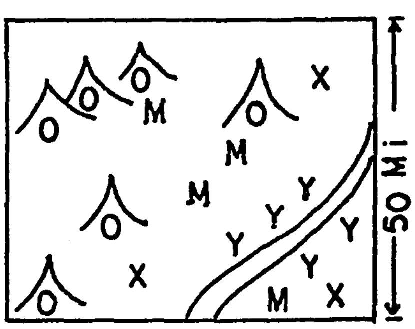
1. "Because it does not have the right number."
2. "Because it has less sugar molecules."
3. "It has more room to shrink."
4. "It has less water and will dry up."

Even though this sample of students made the correct choice to the test item, their reasons did not demonstrate that they had formulated a postulatory-deductive system to provide a rationale for their correct choice. The thought pattern demonstrated by the reasons these subjects gave was one in which they used their senses to identify what was seen and reported this information as a valid explanation for their answer. "Because it does not have the right number and it has less sugar molecules" clearly demonstrate reasons in which the subjects relied on their senses and reported what was real to them. The subject who said, "it has more room to shrink" used a form of intuition, but he incorrectly equated room with concentration thus his intuition revealed a lack of understanding of the meaning of concentration. The fourth response indicates a condition that may have happened under different circumstances but did not relate to concepts of diffusion or concentration. The student who gave this explanation failed to comprehend the meaning of diffusion and concentration. Instead he relied on the meaning that something would dry up if it did not have enough water.

All of the reasons given in the above example were judged incorrect because the subjects neither expressed a postulatory-deductive scheme nor gave implications that suggested such a scheme was inferred. Essentially these reasons were expressions of what the students saw or sensed from information contained in the test item and reported as valid explanations. These reasons were obviously not sufficient to explain responses on formal concept items. On the other hand, these kinds of explanations were judged correct for explaining responses on concrete test items. In other words, explanations for responses on concrete concept items did not require implications derived from a hypothetico-deductive scheme but do require the kinds of reasons found in the next example.

The following item is presented to show that reporting data is sufficient for explaining responses made on concrete test items.

> A 2500-square-mile area was mapped as follows. Four kinds of plants are shown by the letters M,O,X, and Y.



"Which plant appears to require the most moisture?"

The principal concept contained in this item is the relationship between where the plants are located and if this location provides moisture. The subjects had to reason that near the stream means moisture and then locate those plants that grew there. This kind of reasoning depends on first-hand experiences, and the subjects can respond to this test question by observing and reporting concrete data contained in the diagram.

The following reasons given to the above question were judged correct.

1. "More Ys along the river."
2. "Y is surrounding the brook."
3. "Y, because they grow nearer the stream."

The reasons that these students gave indicated that they recognized the relationship between where the plants grew and the ability of this location to supply moisture and picked the appropriate plant (Y) which was found by the stream.

This test item may appear trivial, but there were students who could not follow the pattern of thought found in the above explanations. The following reasons given to the same test item were judged incorrect.

1. "M because it is in the mountain where there is more rain fall."
2. "Because there are more Os in the mountains."
3. "Because there are very few of the Xs."
4. "Because Os are not by the water and Os need it."

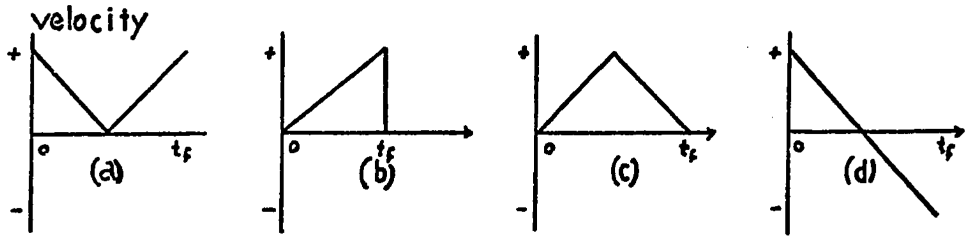
The students who gave these reasons obviously did not understand the relationship that near a stream equals moisture. These reasons demonstrate that the students' explanations were based on information not supported by fact or relationships derived from faulty experiences. For example, the student who picked the plant that grew in the mountain and stated as fact, "... there is more rainfall" gave information not supported by evidence from the test item. The subjects who gave reasons two and three reported facts, "... there are few of them and there are more of them," but the number of these plants at those locations did not meet the condition for moisture requirement. The student who gave reason number four made the statement that, "...

Os need it" which was based on intuition not supported by evidence.

All of the reasons given in the above example with the exception of number four reported information observed from the diagram of the map. The data, however, that were reported did not explain a valid relationship between location and fulfillment of moisture requirement. The reasons, therefore, were judged incorrect and were considered explanations that did not reflect an understanding of the concrete concept found in this test item.

The following explanations given to responses on a physics test item were evaluated by the investigator using the same criteria. The test item used is:

> Which graph below is closest to the velocity versus time graph of a stone which is thrown straight up into the air at time t=0 and returns to earth at $t=t_f$?



The above item contains two concepts that require different levels of thought to derive their meaning. The concept of motion as it is used in this question is well within the ability level of concrete operational thinkers because it is based on first-hand experience. That is, most students have observed the motion of a stone being thrown up and watched it return to earth. The application of the experience needed to answer this test item, however, required transposing this action to a velocity-time graph and arriving at the acceleration curve of gravity. This is an operation on an operation which requires the construction of a hypothetico-deductive scheme within which the concept of gravity derives its full meaning for freely-falling objects. This item, therefore, was rated formal operational.

These reasons were judged correct for responses to the above physics item.

1. "The stone undergoes constant negative acceleration, (d)."
2. "The velocity will immediately decrease at a constant rate and then decrease in a negative direction, (d)."
3. "While in flight, the stone would slow down, reach zero velocity and then begin a negative velocity shown by graph (d)."

The reasons submitted by those students whose explanations were judged correct demonstrated how the motion of the stone and the effect of gravity are related. The explanations stated in effect; _if_ the motion is up and then down, _then_ the velocity is constantly changing due to the effect of gravity, _therefore_, the graphic representation of this motion is a negative slope which corresponds to graph (d). This form of reasoning illustrates that the subjects constructed a hypothetico-deductive scheme and used it to find an implication that could explain and identify the correct response.

The following reasons were judged incorrect.

1. "In the middle of the flight it reaches its peak and gradually reaches zero distance, (c)."
2. "Has high initial velocity and slows down, then while falling returns to initial velocity, (a)."
3. "The velocity would start out high, approach zero as it neared the top of its arch and speed up again as it came down, (a)."

These reasons were judged incorrect because the explanations revealed that the subjects neither understood the implication of gravity on the motion of the stone nor the meaning of a velocity-time graph. In the first explanation the subject implied that the motion of the stone and its trajectory were equivalent to the acceleration curve. The examinee stated, "in the middle of the flight it reaches its peak..." which indicates that what he thought and what he sensed were both based on concrete reality. In the second and third explanations the subjects simply reported a visual description of the stone's velocity as revealed from graph (a). The implication drawn from these explanations is that the thinkers could not think beyond the actual path of the stone or its velocity at various points along the trajectory. These thoughts are tied to the concrete data found in the test item, and therefore, producted explanations that were judged incorrect for formal test items.

A Posteriori Categorization of Students' Explanations and Comparison with Levels of Achievement on One Piagetian Task

In the last section examples of reasons students gave in explaining their responses to the items in the multiple-choice test were presented. Using those examples a demonstration why certain explanations were judged correct or incorrect was provided. The decision as to whether or not a reason was correct was dependent upon the students' understanding of the concepts contained in the test items and whether or not they were able to reach a valid conclusion by either constructing a hypothetico-deductive scheme or applying concrete data found in the test items. The number of correct explanations reported in Tables 9 and 10 was found by using the same criteria and method of evaluation that were used in the previous examples. The data were statistically treated and used to test the principal hypotheses of this study.

The data presented in this section were not intended as part of this investigation. The question of whether or not the reasons students gave to explain their responses to the multiple-choice examinations could be categorized into levels of explanations and compared to the levels of achievements found on the Piagetian tasks was considered a posteriori. Whether or not these categories of reasons represent the same genetic sequences as described by Piaget in his theory of intellectual development remains to be investigated. The data presented represent only the investigator's subjective classification of the explanations found in this study. A description of the categories follows:

<u>Category N</u>: No explanation or statement, "I guessed."[9]

---

[9] All of the statements given in the categories were taken from the biology test item found in <u>Examples of Statements</u>, p.43. The question asked is, "Which cell(s) will shrink?"

Category I: Reasons in which the student gave data that did not apply to the question contained in the test item, e.g., "be-because it is in the middle."

Category II: Reasons in which students gave concrete data as proof of explanations rather than expressing an implication. The data given were related to the question asked in the test item, e.g., "because the cell had the right numbers."

Category III: Reasons in which the students gave implications not tied to concrete data but did not specifically express the relationship between the data in which the implications were inferred. The question of whether or not the implication were correct was not considered; only that the reasons inferred an implication, e.g., "as the water moves out of the cell, its volume and size will shrink."

Category IV: Reasons in which the students verbalized a relationship between the data and gave an implication. The implication may have been incorrect due to faulty relationships between the data and events, but the reasons were evidence that the students used a hypothetico-deductive scheme, e.g., "only cell one has a smaller concentration inside, water moves to a higher concentration, therefore cell one will shrink."

The sample of subjects used for this a posteriori investigation was obtained by assigning a number to each subject in the entire population of biology and physics samples, and randomly selecting 30 subjects from each sample.[10] The reasons submitted by these two subsamples for responses on the 10 formal test items in each subject content examination were categorized according to the classification scheme presented above.

The reason that concrete test items could not be used was because the explanations found in categories III and IV would not be required for students to explain their responses to these test items. Since formal test items do require the kinds of reasons found in all of the categories, the 10 formal items from each subject matter test were used to sample the reasons and classify them into the four categories.

---

[10]The procedure was to pick the first, third and fourth digit to entries in the second and fourth lines on the third and fifth columns in a table of random numbers until 30 subjects in each sample were selected. The table used was taken from C.R.C. Standard Mathematical Tables, op. cit., pp. 237-243.

Tables 14 and 15 show the results of the categorization of reasons for responses on the 10 formal test items.

TABLE 14

FREQUENCY OF CATEGORIES OF REASONS GIVEN ON FORMAL TEST ITEMS: BIOLOGY SAMPLE

Category/ Item	2	3	4	7	9	13	14	15	11	20	%
N	4	3	0	6	7	4	11	8	11	6	20
I	1	0	2	1	3	1	2	2	0	1	4
II	10	11	17	11	12	12	8	11	4	9	35
III	7	12	6	9	6	8	8	6	8	8	25
IV	8	5	5	3	2	5	4	3	6	6	16

TABLE 15

FREQUENCY OF CATEGORIES OF REASONS GIVEN ON FORMAL TEST ITEMS: PHYSICS SAMPLE

Category/Item	6	9	10	11	13	14	15	16	18	20	%
N	4	3	10	3	1	6	4	5	5	8	16
I	2	3	0	0	1	0	2	0	0	0	3
II	9	9	2	5	6	7	8	5	7	6	21
III	7	6	3	8	10	6	8	8	11	4	24
IV	8	9	15	14	12	11	8	12	6	12	36

The frequency of achievement levels on one Piagetian task was obtained by reviewing the data accumulated when the subjects' level of mental development were assessed. Only those responses on the separation-of-variables task were used for making the comparison with the categorization of reasons. This task was selected because it was considered to be less content oriented than the others and therefore not as likely to be influenced by prior knowledge of material contained in the task. The entire populations of the biology and physics samples were used to obtain the number of responses on the Piagetian task. The results are reported in Table 16.

TABLE 16

FREQUENCY OF ACHIEVEMENT LEVELS ON SEPARATION OF VARIABLES TASK

Sample	IIA	IIB	IIIA	IIIB
Biology	5 (6)[a]	41 (46)	25 (28)	18 (20)
Physics	1 (1)	14 (24)	19 (32)	25 (42)

[a] Number in parenthesis are per cents of total sample

Histograms of the categorization of explanations were drawn for each test item[11] and a composite of all ten test items. The histograms of the composites of the ten test items and per cents of achievement levels on the separation-of-variables task for each subject sample were drawn on the same page in order to compare the shapes of the two graphs. Unless designated otherwise, the vertical and horizontal axes represent the frequency of responses and levels of responses respectively.

[11]The formal items on the biology examination are 2, 3, 4, 7, 9, 11, 13, 14, 15, 17 and 20. The formal items on the physics examination are 6, 9, 10, 11, 13, 14, 15, 16, 18, and 20.

FIGURES 1-6

HISTOGRAMS OF FORMAL ITEMS FOR BIOLOGY SAMPLE

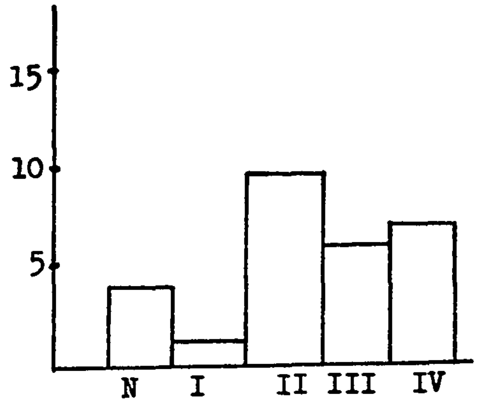


Figure 1: Item 2

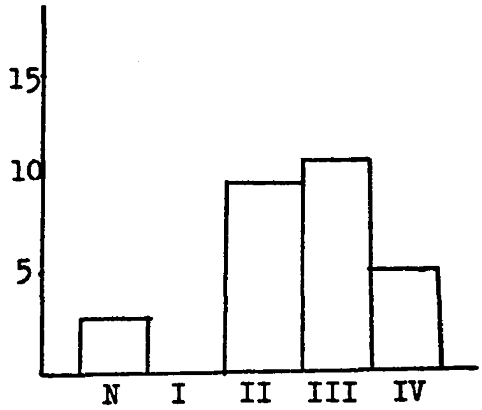


Figure 2: Item 3

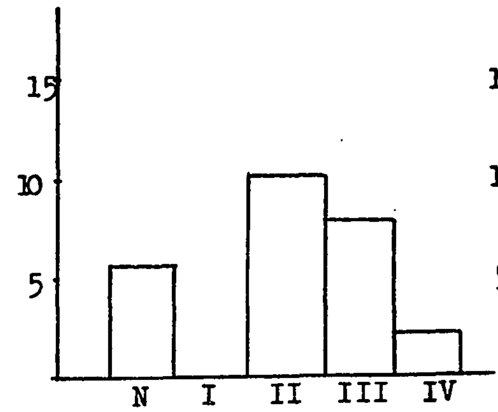


Figure 3: Item 7

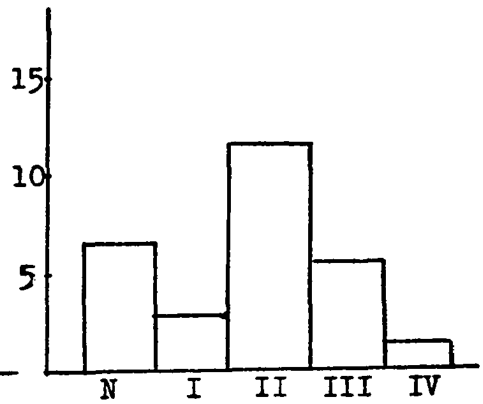


Figure 4: Item 9

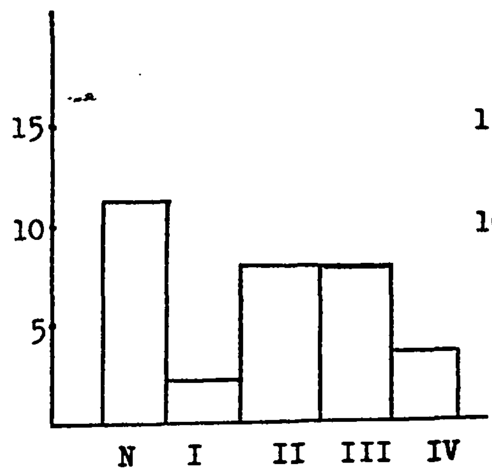


Figure 5: Item 14

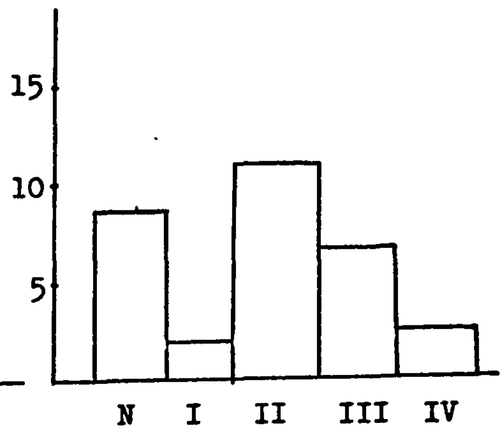


Figure 6: Item 15

FIGURES 7-10

HISTOGRAMS OF FORMAL ITEMS FOR BIOLOGY SAMPLE (continued)

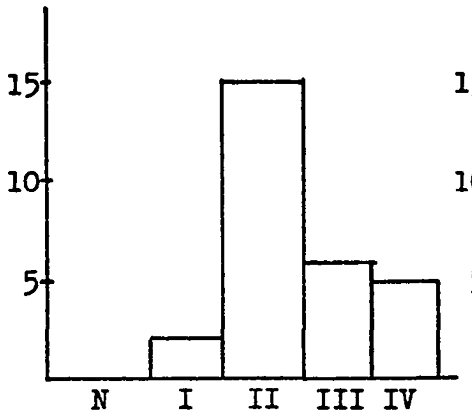


Figure 7: Item 4

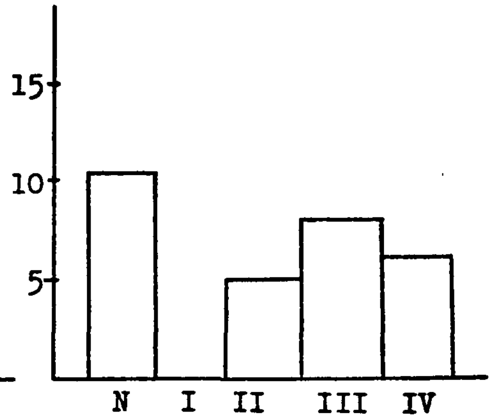


Figure 8: Item 11

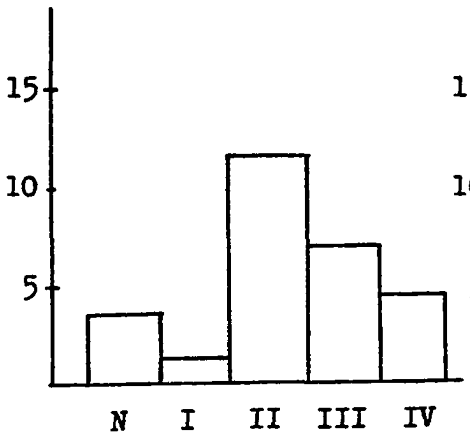


Figure 9: Item 13

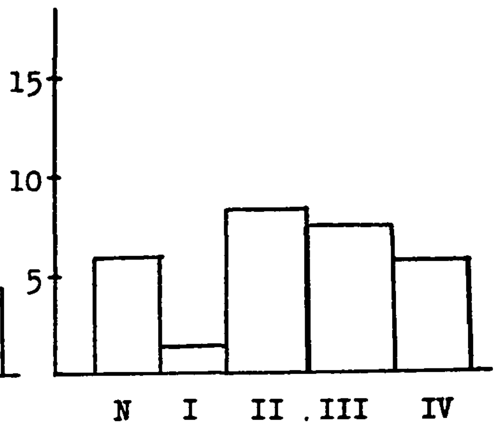
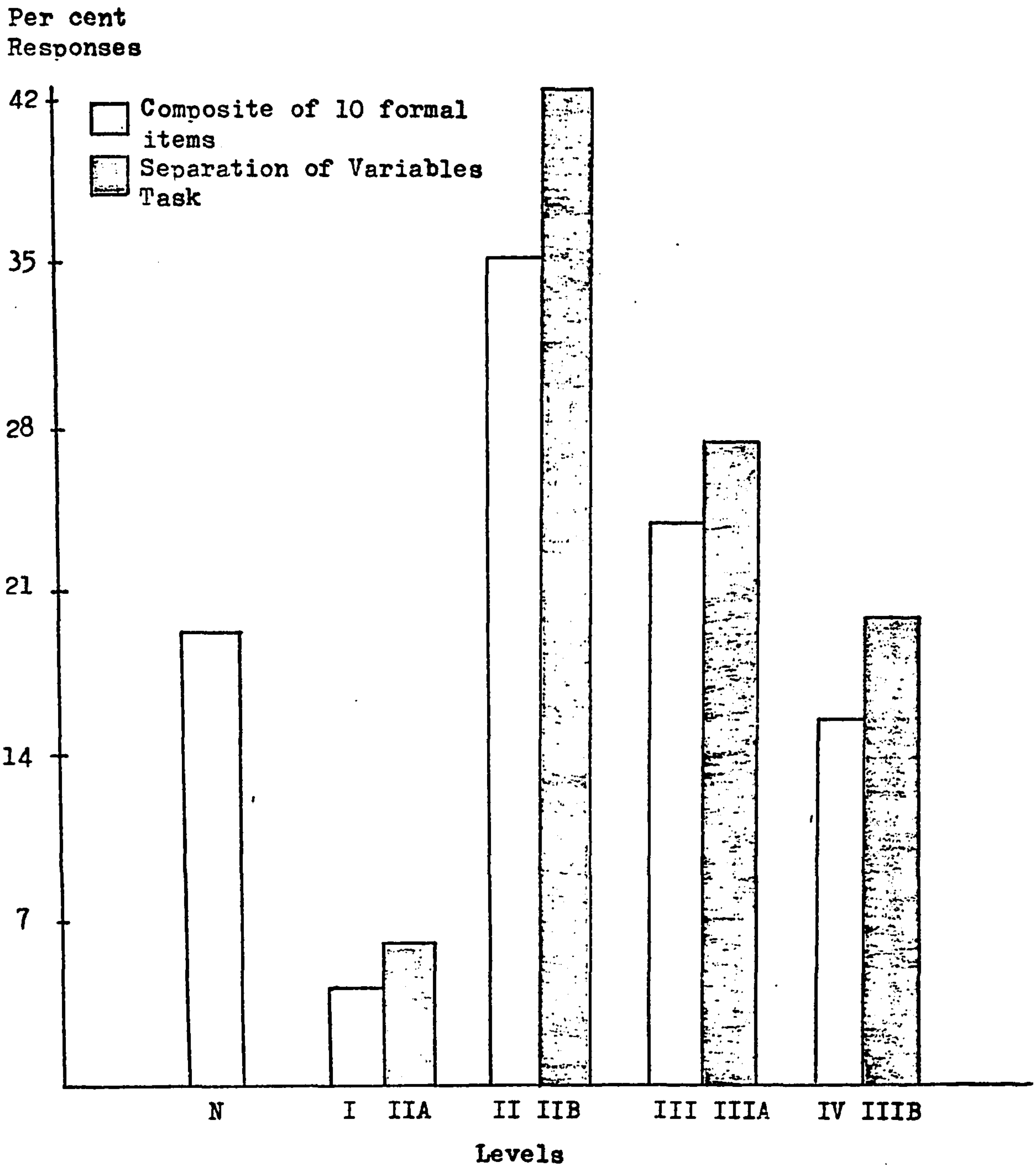


Figure 10: Item 20

FIGURE 11

A COMPARISON OF CATEGORY OF RESPONSES ON FORMAL ITEMS WITH LEVELS OF ACHIEVEMENT ON SEPARATION OF VARIABLES

Biology Sample



FIGURES 12-17

HISTOGRAMS OF FORMAL ITEMS FOR PHYSICS SAMPLE

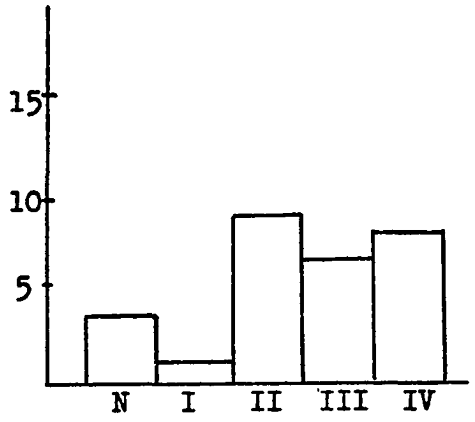


Figure 12: Item 6

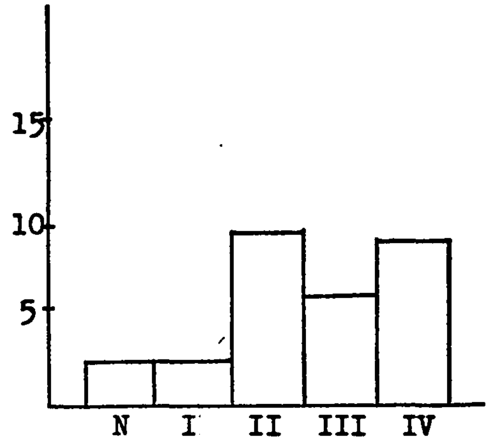


Figure 13: Item 9

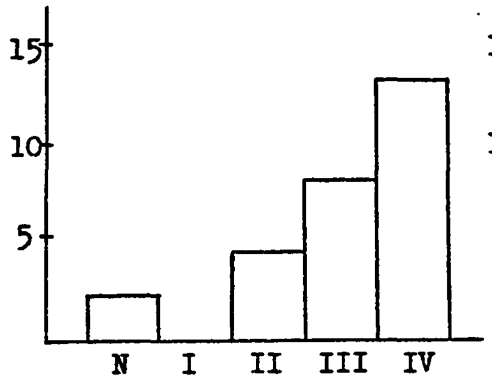


Figure 14: Item 11

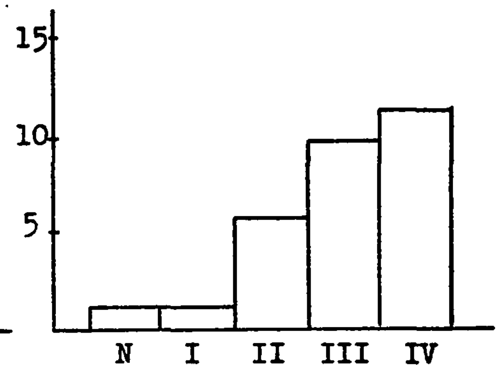


Figure 15: Item 13

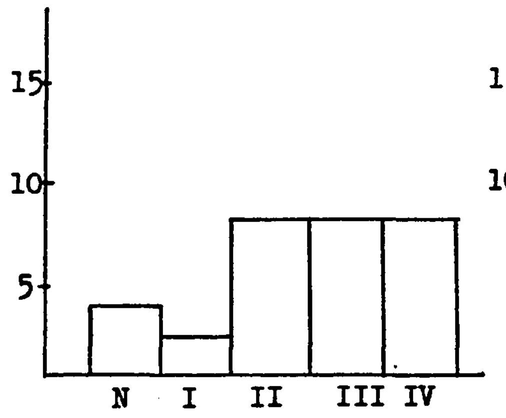


Figure 16: Item 15

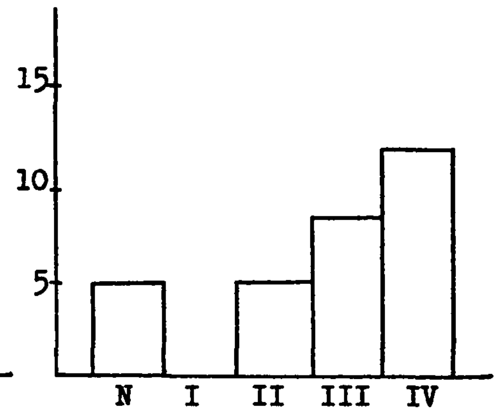


Figure 17: Item 16

FIGURES 18-21

HISTOGRAMS OF FORMAL ITEMS FOR PHYSICS SAMPLE (continued)

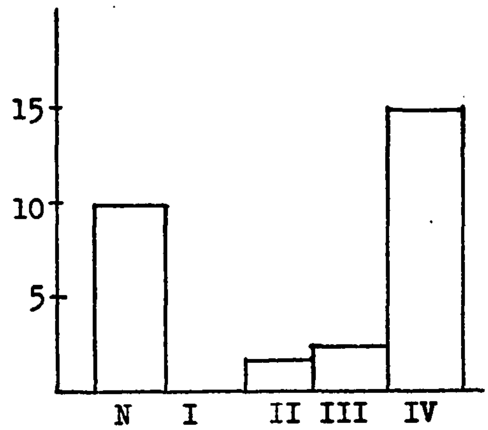


Figure 18: Item 10

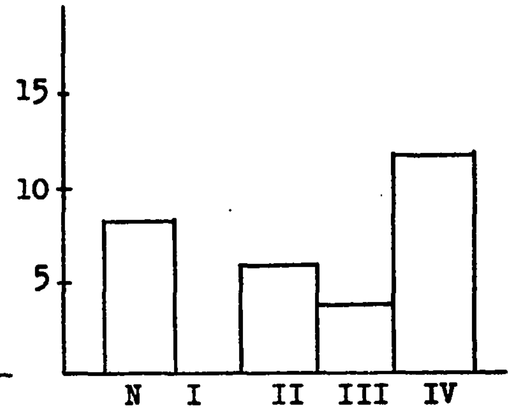


Figure 19: Item 20

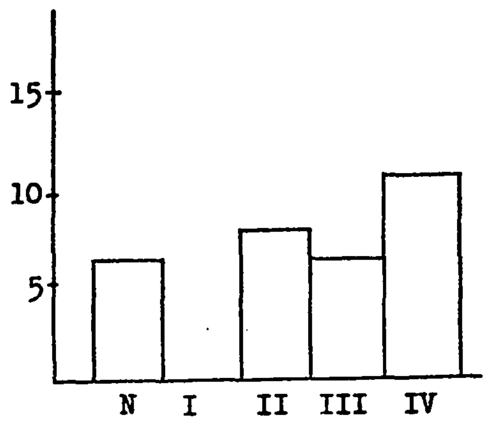


Figure 20: Item 14

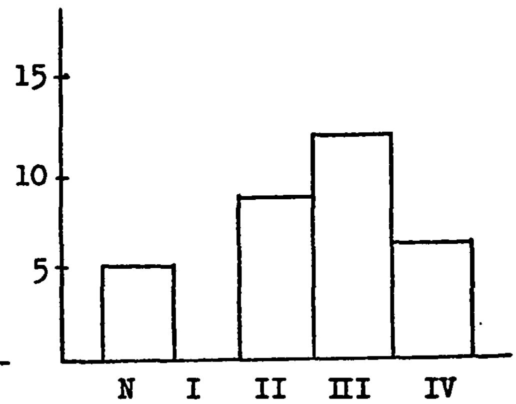
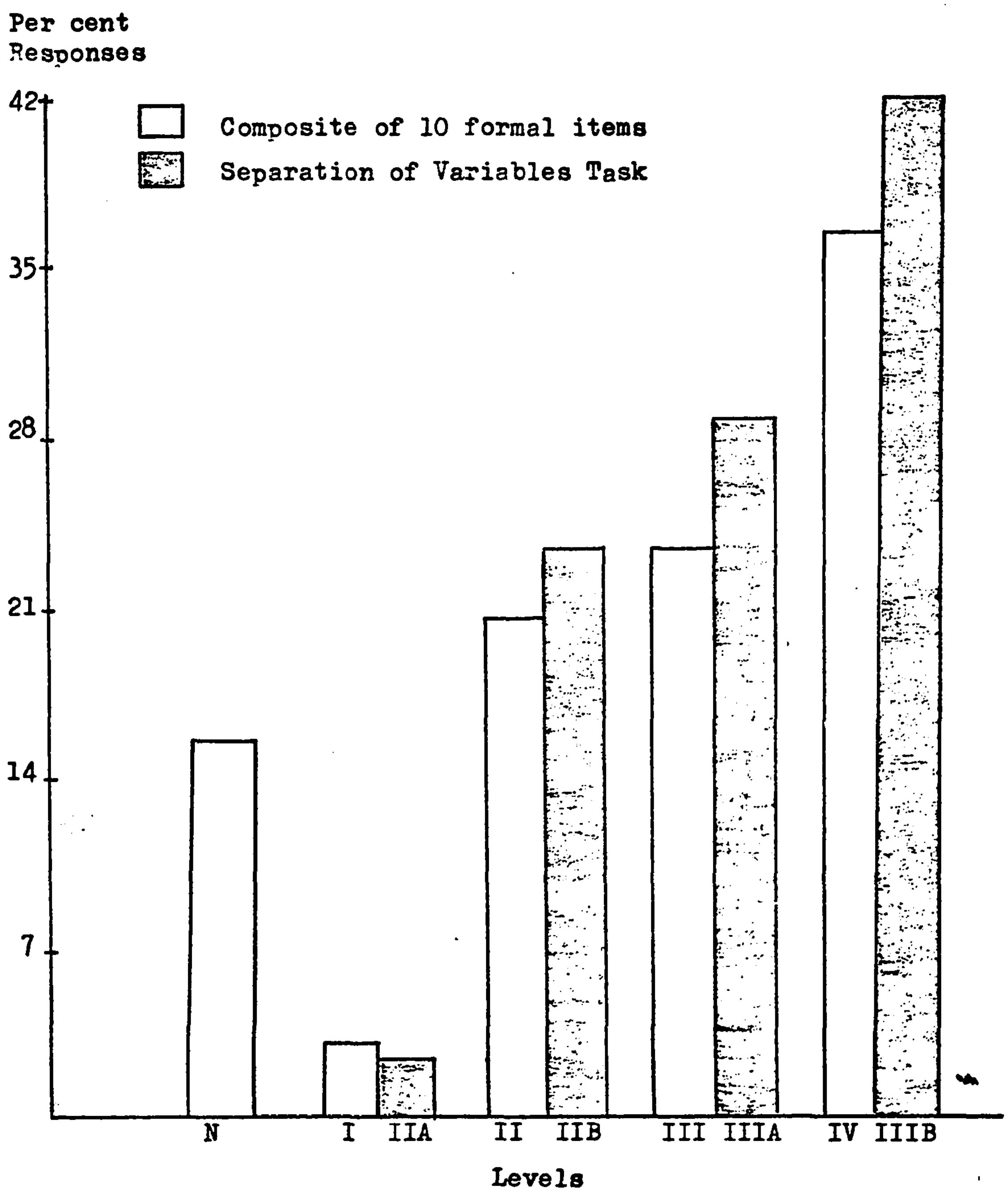


Figure 21: Item 19

FIGURE 22

A COMPARISON OF CATEGORY OF RESPONSES ON FORMAL ITEMS WITH LEVELS OF ACHIEVEMENT ON SEPARATION OF VARIABLES

Physics Sample



Summary of Results

The shapes of the histograms for the composite of all ten formal items and the levels of achievement of the separation-of-variables task for each sample were quite similar. The only difference appears in the Category N portion of the graphs. Since the levels of achievement on the Piagetian task did not allow for a "no-response" category, a comparison of this category was not possible. Otherwise the relative per cents of responses in each category and levels of achievement on the Piagetian task did not vary more than 11% for each sample. This deviation of per cents for each category of explanation and levels of achievement can be seen from the following table.

TABLE 17

COMPARISON OF PER CENT OF RESPONSES IN EACH CATEGORY AND LEVEL OF ACHIEVEMENT

Biology Sample

Category/Level	N	I/IIA	II/IIB	III/IIIA	IV/IIIB
10 Formal Items	20	4	35	25	16
Separation-of-Variables Task		6	46	28	20
Deviation	--	± 2	± 11	± 3	± 4

Physics Sample

Category/Level	N	I/IIA	II/IIB	III/IIIA	IV/IIIB
10 Formal Items	16	3	21	24	36
Separation-of-Variables Task		2	24	32	42
Deviation	--	± 1	± 3	± 8	± 6

The comparison of responses in each category and level of achievement found in Table 17 also reveals which level had the highest per cent of responses. In the biology sample the highest per cent of responses fell in category II (35%) and level IIB (46%). The similarity of these two frequencies of responses would indicate that category II explanations were characteristic of concrete operational thinkers. Furthermore, the higher frequency of the level would support the claim that most of the students in biology are concrete operational.[12]

The results of the physics sample show a higher per cent of IIIB responses (42%) and category IV explanations (36%). Since IIIB is a formal operational level of achievement, this would indicate that category IV was the kind of explanation expected of formal operational thinkers. Furthermore, the frequency of responses at these two levels (category IV and IIIB) would support the claim that most of the students in high school physics are operation at the formal operational level.[13]

In sum, the results of this a posteriori investigations would indicate that categorization of written explanations for responses on formal test items are possible and appear to have educational merit. The explanations students give can be readily classified into levels according to <u>what</u> information in the test questions is used and <u>how</u> it is applied to explain their responses. By utilyzing the categories of levels of explanations, a first approximation for identifying formal and concrete operational thinkers can be made. It should be pointed out that these categories do not necessarily represent the same genetic sequences of intellectual development described by Piaget. The categorization of written explanations may, however, provide an alternate method of identifying levels of mental development using non-Piagetian tasks.

---

[12]For this study 63% of the subjects in the biology sample was found to be concrete operational.

[13]The per cent of physics subjects in this study was found to be 66% formal operational.

CHAPTER IV

CONCLUSIONS, RECOMMENDATIONS AND QUESTIONS FOR FURTHER STUDY

The primary objective of this study was to determine if a student's responses on multiple-choice items reflect understanding of formal and concrete concepts and if that understanding was commensurate with the student's level of mental development. The problem was to find the relationship among the reasons given for selecting a particular response, the level of concepte used in the items, and the thinking levels of the students. These relationships then would determine which concepts could be understood by formal and concrete operational thinkers.

Three null hypotheses were formulated and tested. Table 18 lists each hypothesis, the results of the statistical treatment of the data found, and the decision made.

TABLE 18

SUMMARY OF RESULTS AND DECISION MADE FOR EACH HYPOTHESIS

Ho_1:	There is no correlation between the scores made on the multiple-choice test and the number of correct reasons given for formal and concrete test items by formal operational subjects.
Results:	Formal operational students were able to correctly respond to both formal and concrete test items by giving a significant number of valid reasons that could explain their answers.
Decision:	Reject the null hypothesis and accept the alternate hypothesis.[a]
Ho_2:	There is no correlation between the scores made on the multiple-choice test and the number of correct reasons given for formal test items by concrete operational subjects.
Results:	Concrete operational students were not able to respond correctly to formal concept items as demonstrated by a significant number of incorrect or no reasons given to these items.
Decision:	Accept the null hypothesis.
Ho_3:	There is no correlation between the scores made on the multiple-choice test and the number of correct reasons given for concrete test items by concrete operational subjects.
Results:	Concrete operational students were able to respond correctly to concrete items and were able to give a significant number of correct reasons that explained their responses.
Decision:	Reject the null hypothesis and accept the alternate hypothesis.[b]

[a]The alternate hypothesis states that there is a positive correlation ... for formal and concrete items ... by formal operational subjects.

[b]The alternate hypothesis states that there is a positive correlational ... for concrete items ... by concrete operational subjects.

Conclusions

The following conclusions are made as a result of the findings of this study:

1. Those biology and physics subjects who were found to be concrete operational do not exhibit the capacity to understand formal concepts. These students were unable to develop a hypothetico-deductive scheme that could be used to correctly explain their responses to formal test items. Their responses to these test items revealed either a dependency on present reality as they reported concrete data given in the test item as proof of their explanations or that they guessed the answers and could not give a written explanation.

2. Concrete operational thinkers in both samples are able to understand the meaning of concrete concepts. The significant positive correlation between scores and correct reasons given to concrete test items revealed that concrete operational thinkers possessed ideas that enabled them to develop the meaning of concrete concepts.

3. Formal operational thinkers in both samples possess the mental abilities to understand the meaning of abstract or formal and concrete concepts. The reasons given for responses on formal and concrete test items showed that these thinkers were able to express their ideas in an if ... then ... therefore ... reasoning scheme to reach valid conclusions.

4. Objective tests can provide a valid measure of understanding of formal and concrete concepts if the operational level of each student is known. Those students who can substantiate their responses to formal and concrete concept items are operating at the formal level and the test scores are a valid measure of understanding the meaning of these concepts. Objective test scores of concrete operational thinkers do not reflect understanding of formal concepts because these thinkers can not give valid explanations for their responses. Concrete objective test questions, however, can be used to measure understanding of these concepts by concrete operational thinkers.

Recommendations

This investigator believes that the following general recommendations to high school teachers are warranted.

1. An initial indication of a student's level of mental development can be obtained by using one Piagetian task, the conservation of volume. This task is relatively simple to administer and the responses are easily categorized into concrete and formal operational levels. The results from administering this task early in the school year can provide teachers with information that will help fit the subject matter content to the level of the learner. At the same time, this information can provide teachers with a yardstick for determining which concepts a particular learner can successfully interact with in a testing situation.

2. When objective tests are used, teachers should identify the level of concepts found in the test items. This can be accomplished by answering the following questions for each test item.

a) Does the question ask for concrete data that can be readily sensed from information given or developed from first-hand experience? If so, this is a concrete concept item and one which concrete operational thinkers can successfully respond to.

b). Does the question require the construction of a postulatory-deductive scheme in order to draw inferences and make generalizations? This is a formal concept item and one with which formal operational thinkers can interact.

The following recommendations are made to biology and physics teachers who plan to use objective examinations to measure concept understand.

1. Since the majority of students in the biology sample of this study was found to be concrete operational, examinations written for these thinkers should contain many more concrete concept items than formal.

2. Examinations written for physics students should contain more formal concept items than concrete because physics teachers can expect to find approximately 65% of their students formal operational.

3. The length of the objective examinations for both subject areas should be constructed so that the students do not have to respond to more than 20 items.

4. When formal concept items are used in the objective tests, the students should be asked to give a written explanation for each of their responses to these items. The number of items in the examinations should be reduced to less than 20 items if this is done.

A final recommendation is made to biology teachers concerning the level of content found in most textbooks used in high school. The majority of material found in Biological Science Curriculum Study texts and other current textbooks reviewed by this investigator contain formal concepts. These concepts are inappropriate for concrete operational thinkers. If these concrete learners are to be successful in developing understanding of formal concepts, teachers must present the material in the proper sequence in order to help lead these learners from concrete to formal thinking patterns.

Suggestions for Further Study

The following hypothesized guidelines for judging the operational level of thinkers based on their written explanations to formal test items need to be investigated.

1. Students who do not give a written explanation or state that they guessed are operating at the pre-concrete level.

2. Students who report concrete data from the test items as proof of their explanations are concrete operational.

3. Students who make inferences that lead to correct responses but do not give a complete statement of explanation are beginning formal operational.

4. Students who are able to give explanations that show implications based on a postulatory-deductive scheme and to verbalize this scheme are formal operational.

The test of significance between positive correlations of biology and physics samples of this study was performed and reported: see, Table 13, p. 39. An inspection of those positive correlations within each sample (Tables 11 and 12), based on the results of the test of significance performed, shows that the difference between the correlations within the biology sample (.82 and .88) is not significant. The difference between the correlations of subjects within the physics sample (.79 for concrete thinkers and .94 for formal thinkers), however, prove to be significant at the .05 level of confidence. Since both of these correlations were between test scores on concrete items and reasons given by concrete and formal operational subjects and were found to be positive at the .005 level, this would imply that formal operational thinking facilitates the understanding of concrete concepts. An in depth study, therefore, needs to be conducted between those positive correlations reported within the physics sample of this study. This study could determine whether or not formal thought enables formal operational thinkers to understand concrete concepts from a different frame-of-reference than concrete operational thinkers.

Another suggestion for further study arises from the question as to why some concrete operational students (not a significant number) were able to make correct responses to concrete concept items but could not give a written explanation. This was also found to be true for formal operational subjects responding to formal test items. According to this study, concrete thinkers and formal thinkers are able respectively to understand formal and concrete concepts. Apparently, the ability of some formal thinkers to respond to formal items and some concrete thinkers to respond to concrete test items and the ability of each to express in writing a valid explanation for his choice are not the same mental operation for all formal and concrete operational students. The fact that expository writing may be a formal operation obviously not completely developed by concrete thinkers and the development of this process possibly impeded for some formal thinkers needs to be explored.

The question as to whether or not formal and concrete concepts can be identified in non-science subject areas needs to be answered. Answers to this question can provide information that may bring these subjects in line with the progress made in the identification and application of concepts in scientific subject content.

BIBLIOGRAPHY

Books

Biological Science Curriculum Study, Biological Science: An Inquiry into Life, Yellow Version, New York: Harcourt Brace, 1968.

Ferguson, George A., Statistical Analysis in Phychology and Education, New York, McGraw-Hill, 1966.

Ginsbery, Herbert and Opper, Sylvia, Piaget's Theory of Intellectual Development, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1969.

Hoffman, Banesh, The Tyranny of Testing, New York: Collier Books, 1964.

Inhelder, Bärbel and Piaget, Jean, The Growth of Logical Thinking from Childhood to Adolescence, New York: Basic Books, Inc., 1958.

Kerlinger, Fred N., Foundations of Behavioral Research, New York: Holt, Rinehart and Winston, Inc., 1964.

Lindquist, E.F., Editor, Educational Measurements, Washington, D.C.: American Council on Education, 1951.

Phillips, John L., Jr., The Origins of Intellect: Piaget's Theory, San Francisco: W.H. Freeman and Co., 1969.

Piaget, Jean, The Psychology of Intelligence, Paterson, New Jersey: Littlefield, Adams, 1963.

Renner, John W., Bibens, Robert F., and Sheperd, Gene D., Guiding Learning in the Secondary School, New York:

Renner, John W., and Stafford, Don G., Teaching Science in the Secondary School, New York: Harper and Row Publishers, 1972.

Periodicals

Karplus, Robert and Karplus, Elizabeth F., "Intellectual Development Beyond Elementary School I: Deductive Logic," School Science and Mathematics, May, 1970.

Karplus, Robert and Peterson, Rita W., "Intellectual Development Beyond Elementary School II: Ratio, A Survey," School Science and Mathematics, December, 1970.

Kohlberg, Lawrence and Gilligan, Carol, "The Adolescent as a Philosopher: The Discovery of the Self in a Postconventional World," Daedlus, Fall, 1971, 100.

Lawson, Anton E., and Renner, John W., "A Quantitative Analysis of Responses to Piagetian Tasks and Its Implications for Curriculum," Science Education, Vol, 58, No. 3., (In Press).

Piaget, Jean, "Development and Learning," Journal of Research in Science Teaching, 1964, 2, No.3.

Piaget, Jean, "Intellectual Evolution from Adolescence to Adulthood," Human Development, 15: 1-12, 1972, No. 1.

Unpublished Materials

Friot, Faith E., The Relationship Between an Inquiry Teaching Approach and Intellectual Development, (Unpublished Doctoral Dissertation), The University of Oklahoma, 1970.

Lawson, Anton E., An Investigation into Secondary School Science Curricular and the Intellectual Level of the Learner, (Unpublished Doctoral Dissertation), The University of Oklahoma, 1973.

APPENDIX I

Subject Matter Examinations

BIOLOGY EXAMINATION

Directions: Select the letter that corresponds to the best answer to the test item and mark it on the answer sheet. Give a written reason(s) in the space below the item that can explain your choice. If you are unable to give a reason, but think you know the answer, mark the answer sheet. This examination will not effect your grade in biology.

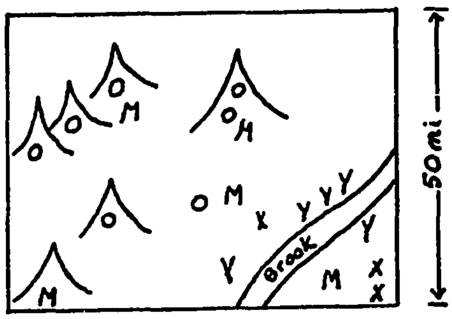
1. What magnification would let you see the most paramecia at one time?

a) 50x b) 100x c) 200x d) 400x

reason: (concrete 73 .62)[1]

The following four questions are based on the information below.

A 2500-square-mile area was mapped as follows. Four kinds of plants are shown by the letters M, O, X, and Y .



2. If each M represents a single plant, what is the density of M in the area?

a) 1 per 10 square-miles b) 1 per 25 square-miles

c) 1 per 50 square-miles d) 1 per 500 square-miles

reason: (formal 39 .61)

3. If there were only a total of 250 plants in the area, the density of these plants would be most nearly

a) 1 per square-mile. b) 5 per square-mile.

c) 1 for every 5 square-mile. d) 1 for every 10 square-mile.

reason: (formal 53 .29)

4. Which plant seems best adapted to conditions of the entire area?
a) M b) O c) X d) Y

reason: (formal 32 .61)

---

[1] Information located here and below indicates; level of concept, difficulty index, and internal-consistency index.

5. Which plant appears to require the most moisture?

a) M b) O c) X d) Y

reason: (concrete 84 .20)

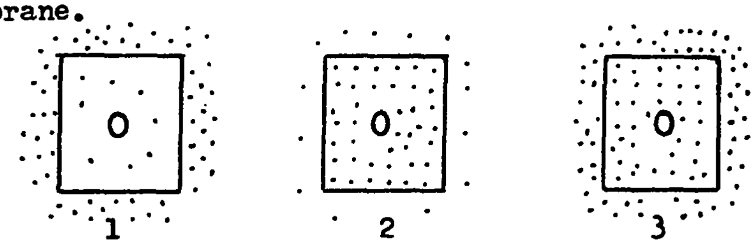
6. What is the area of a plot 60 feet by 60 feet?

a) 120 ft^2 b) 360 ft^2 c) 1200 ft^2 d) 3600 ft^2

reason: (concrete 69 .59)

The next five questions are based on the following information and diagrams of 3 cells in a water-sugar solution.

The dots show sugar molecules which cannot pass through the cell membrane.



7. Into which cell(s) will the most water molecules diffuse?

a) 1 only b) 2 only c) 1 and 2 only d) 1, 2 and 3

reason: (formal 38 .55)

8. In which cell(s) would there be an equal exchange of water molecules?

a) 1 only b) 2 only c) 3 only d) all of them

reason: (concrete 68 .18)

9. Which cell(s) will shrink in size?

a) 1 only b) 2 only c) 3 only d) all of them

reason: (formal 43 .31)

10. In which cell(s) will no change in size occur?

a) 1 only b) 2 only c) 3 only d) 1 and 2

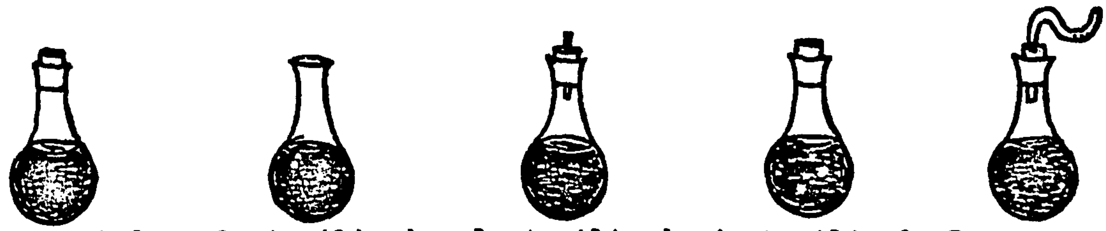
reason: (concrete 63 .10)

11. If the dots were salt molecules that could move through the cell membranes, the most salt would move out into the water through cell(s)

a) 1 only. b) 2 only. c) 3 only. d) 1 and 2.

reason: (formal 57 .52)

The next four questions are based on the following experiment with flasks of broth treated in different ways.



1 untreated 2 sterilized 3 sterilized 4 sterilized 5 sterilized

12. Flask 1 serves as a direct control for flask

a) 2. b) 3. c) 4. d) 5.

reason: (concrete 53 .58)

13. If organisms in the broth are killed by sterilization, they will appear first in flask

a) 1. b) 2. c) 3. d) 5.

reason: (formal 36 .70)

14. If no organisms appeared in flasks 4 or 5, the idea of spontaneous generation would be

a) supported. b) unsupported. c) not affected. d) proved.

reason: (formal 52 .58)

15. The hypothesis that all life comes from existing life would be supported if no organisms appeared in flasks

a) 1 and 2. b) 2 and 3. c) 3 and 4. d) 4 and 5.

reason: (formal, 43 .32)

The next four questions are based on the hypothesis that gland X, Y, and Z influences one another. They may do so as follows:

1. $X \rightarrow Y \rightarrow Z \rightarrow$

2. $$\begin{tikzcd} & Y \arrow[d] \\ X \arrow[ur] \arrow[r] & Z \end{tikzcd}$$

3. $Z \rightarrow Y \rightarrow X \rightarrow$

4. $Y \rightarrow X \rightarrow X \rightarrow$

16. When gland X is removed, first gland Y stops functioning, then gland Z stops. This suggests pattern

a) 1. b) 2. c) 3. d) 4.

<u>reason</u>: (concrete 60 .73)

17. Gland Z is removed. Stimulation of Y then causes X to secrete. This suggests pattern

a) 1. b) 2. c) 3. d) 4.

<u>reason</u>: (concrete 63 .60)

18. If Y is removed and X and Z continue to function, this suggests pattern

a) 1. b) 2. c) 3. d) 4.

<u>reason</u>: (concrete 52 .53)

19. If Z is removed and X stops functioning this suggests pattern

a) 1. b) 2. c) 3. d) 4.

<u>reason</u>: (concrete 46 .55)

20. The simplest way to find the number of individuals in a large population is to

a) count all of them. b) take a random sample.

c) count the first 500. d) count all in a small area.

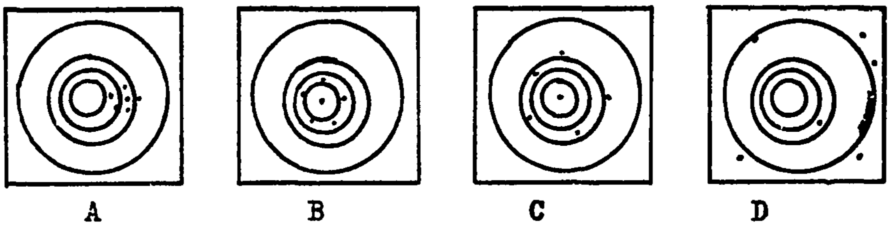
<u>reason</u>: (formal 20 .10)

Key

1.	a	6.	d	11.	b	16.	a
2.	d	7.	b	12.	c	17.	c
3.	d	8.	c	13.	a	18.	b
4.	a	9.	a	14.	b	19.	d
5.	d	10.	c	15.	d	20.	b

PHYSICS EXAMINATION

The following four questions are based on the diagrams below and the following definitions. Accuracy refers to the closeness of a measurement to the accepted value for a specific physical quantity and precision is the agreement among several measurements that have been made in the same way.



1. Six shots were fired at each target. In which target was the accuracy good but the precision poor?

   a) A b) B c) C d) D

reason: (concrete 54 .65)

2. In which target was the precision good but the accuracy poor?

   a) A b) B c) C d) D

reason: (concrete 59 .62)

3. In which target was the precision and accuracy both poor?

   a) A b) B c) C d) D

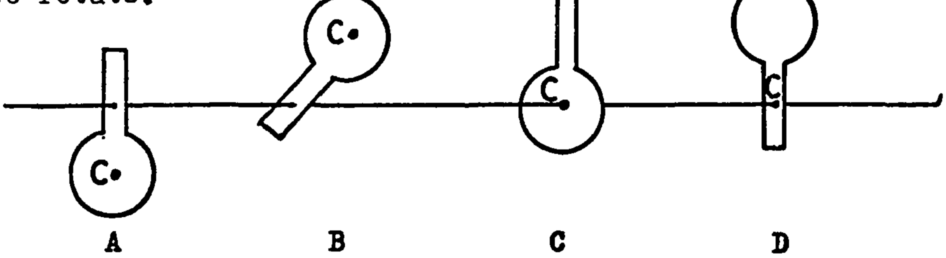
reason: (concrete 90 .78)

4. In which target was the precision and accuracy both good?

   a) A b) B c) C d) D

reason: (concrete 90 .69)

In the diagrams below the objects are suspended by the wire and are free to rotate.



5. If c is the center of gravity of the objects, which object is the least stable?

a) A b) B c) C d) D

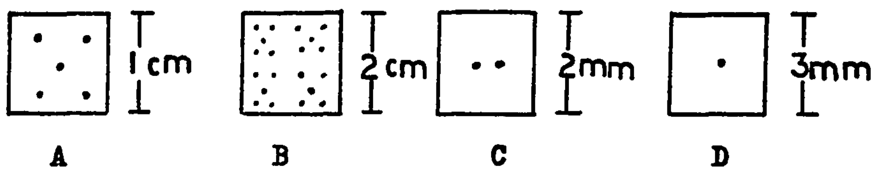
reason: (concrete 89 .72)

6. Which object(s) represent(s) stable equilibrium?

a) A only b) D only c) A and C d) C and D

reason: (formal 32 .45)

7. Which square below has the greatest density of particles?



a) A b) B c) C d) D

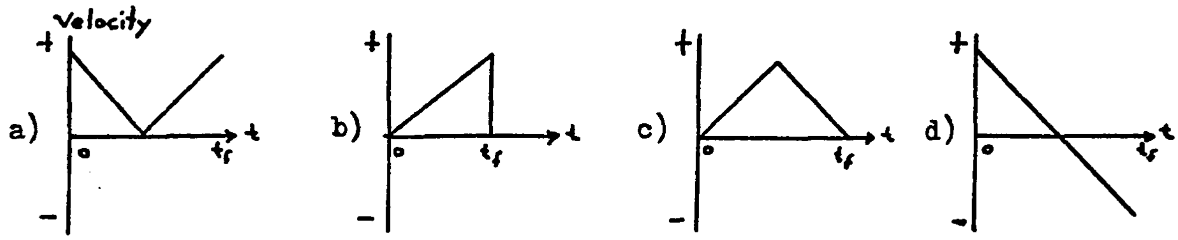
reason: (concrete 50 .61)

8. Which two squares have the same density?

a) A and B b) A and C c) B and D d) C and D

reason: (concrete 72 .49)

9. Which graph below is closest to the velocity versus time graph of a stone which is thrown straight up at time t=0 and returns to earth at $t=t_f$?



reason: (formal 64 .38)

10. What is the speed of the tip of a second hand 2 cm long?

a) $\pi/30$ cm/s b) $\pi/15$ cm/s c) π cm/s d) 2π cm/s

reason: (formal 63 .48)

11. Which vector represents the acceleration of the second hand tip at 12 O'clock?

a) ← b) → c) ↘ d) ↓

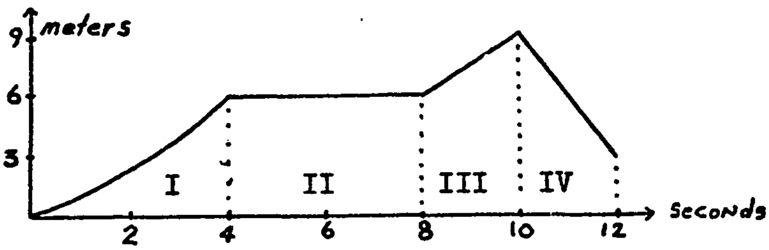
reason: (formal 36 .28)

12. A cart travels 10.0 ± 0.2 cm in 2.0 ± 0.2 seconds. What is the maximum speed of the cart?

a) 4.6 cm/s b) 5.0 cm/s c) 5.4 cm/s d) 5.7 cm/s

reason: (concrete 64 .53)

The motion of a cart is described in the graph below. The next three questions are based on the distance versus time graph.



13. Which portion of the graph indicates the greatest average velocity?

a) I b) II c) III d) IV

reason: (formal 41 .54)

14. Which period(s) during the carts <u>motion</u> does/do it have zero acceleration?

a) I only b) III only c) II and III d) III and IV

reason: (formal 41 .38)

15. If the cart started at rest, what is the final velocity at the end of four seconds?

a) 1.5 m/s b) 2.5 m/s c) 3.0 m/s d) 3.5 m/s

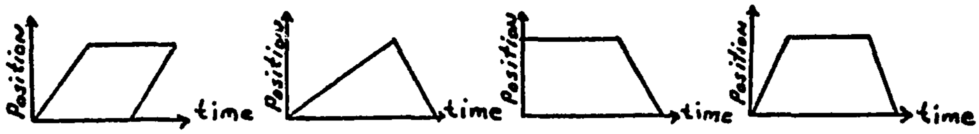
reason: (formal 13 .24)

16. A man drops a ball from a height of 6 ft while riding in an elevator. Which factor will cause the shortest time for the ball to strike the floor of the elevator?

a) elevator is rising at a constant speed
b) elevator is falling at a constant speed
c) elevator is accelerating upward
d) elevator is accelerating downward

reason: (formal 68 .32)

17. Which graph of a car trip below does not represent a real situation?



a) b) c) d)

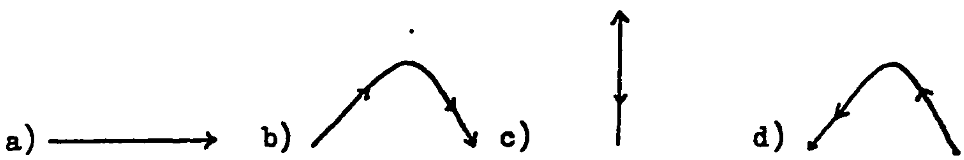
reason: (concrete 95 .77)

18. An arrow shot vertically upward reaches a height of 128 ft during the first second on its way to a higher altitude. What was its initial velocity?

a) 128 ft/s b) 132 ft/s c) 144 ft/s d) 160 ft/s

reason: (formal 24 .37)

19. A ball is thrown vertically upward with a velocity of 24 m/s from a flatcar moving horizontally to the right with a velocity of 4 m/s. Which path of the ball will an observer on the ground see?



reason: (concrete 64 .53)

20. Neglecting wind resistance, what is the distance the ball will travel relative to the ground?

a) zero b) 9.6 m c) 19.2 m d) 24 m

reason: (formal 27 .42)

Key

1.	c	6.	a	11.	d	16.	c
2.	a	7.	c	12.	d	17.	a
3.	d	8.	a	13.	d	18.	c
4.	b	9.	d	14.	d	19.	b
5.	b	10.	b	15.	c	20.	c

APPENDIX II

Panel of Judges

PANEL OF JUDGES

The following persons who had demonstrated a good understanding of both Piagetian theory and the discipline of biology and physics were used to rate the objective test items. Those members who rated the biology items were A.E. Lawson and C.A. Lawson. The physics items were rated by Robert Karplus, Mary Budd Rowe and Don E. Stafford.

In cases where there were disagreement about the level of an item, the investigator served as the final judge of the concept level. If an item was found to be inappropriate or non-functioning by the panel, it was replaced by the investigator with one that was considered suitable and that met the criteria requirements of concept understanding.