DOCUMENT RESUME

ED 157 778 SE-024 780

AUTHOR	Helgeson, Stanley L., Ed.; Blosser, Patricia E., Ed.
TITLE	Investigations in Science Education, Vol. 3, No. 1. Expanded Abstracts and Critical Analyses of Recent Research.
INSTITUTION	Ohio State Univ., Columbus. Center for Science and Mathematics Education.
PUB DATE	77
NOTE	67p.; Contains occasional light and broken type
AVAILABLE FROM	Information Reference Center (ERIC/IRC), The Ohio State Univ., 1200 Chambers Rd., 3rd Floor, Columbus, OH 43212 (Subscription, $6.00; $1.75 ea.)
EDRS PRICE	MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS	*Abstracts; *Cognitive Development; *Educational Research; *Inservice Teacher Education; Instruction; Research; Research Methodology; *Science Education; Teacher Education
IDENTIFIERS	*Piaget (Jean); Research Reports

ABSTRACT

This issue provides analytical abstracts, prepared by science educators, of research reports. Two clusters of research are investigated. The first cluster, Piagetian Studies, contains seven studies which reflect a current trend in science education. The second cluster, Inservice Education, contains two studies which examine the outcomes of inservice education programs. This issue also contains a response to an analysis of a research report. Each abstract includes research purpose, rationale, research design and procedure, findings, interpretations, abstractor's analysis of the research and references. (Author/HM)

INVESTIGATIONS IN SCIENCE EDUCATION

National Association for Research in Science Teaching

ERIC Clearinghouse for Science, Mathematics,
and Environmental Education

Published Quarterly by

The Center for Science and Mathematics Education
College of Education
The Ohio State University
1945 North High Street
Columbus, Ohio 43210

Subscription Price: $6.00 per year. Single Copy Price: $1.75.
Add 25¢ for Canadian mailings and 50¢ for foreign mailings.

TABLE OF CONTENTS

Volume 3, Number 1

TABLE OF CONTENTS (Continued)

NOTES . . .

from the Editor

There are two clusters of research reports in this issue of ISE. The first cluster, PIAGETIAN STUDIES, contains seven studies. The predominance of Piagetian research analyzed in this issue appears to reflect a current trend in science education. The second cluster, INSERVICE EDUCATION, contains two studies which examine the outcomes of inservice education programs.

This issue also contains, for the first time, a response to an analysis of a research report. We believe this kind of exchange is beneficial to the profession and strongly encourage others to send us their publishable responses to critical analyses. Such responses need not be from the author of the research study being analyzed.

Stanley L. Helgeson
Editor

Patricia E. Blosser
Associate Editor

PIAGETIAN STUDIES

Kolodiy, George. "The Cognitive Development of High School and College Science Students." The Journal of College Science Teaching, 5(1): 20-22, September, 1975.
Descriptors--*Cognitive Development; *College Students; Higher Education; *Intellectual Development: Prediction: Science Education; Secondary Education; *Secondary School Science

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Gene F. Craven, Oregon State University.

Purpose

The stated purpose of this study was "to obtain indications of changes in mental level from high school through college and to investigate the relationship of cognitive level to SAT scores, college grades, and attrition." (Kolodiy, 1975; p. 20)

Rationale

Piagetian theory continues to make important contributions to our knowledge of cognitive development. In the past decade, studies of the child's acquisition of knowledge have broadened into formal education. It is only recently, however, that researchers have looked at the relevance of Piagetian theory for high school and college student learning.

Piaget views human cognition as a complex process of interacting with the outside world that includes two simultaneous and complementary aspects which he refers to as assimilation and accommodation. By assimilation he means actively selecting and interpreting external objects and events in terms of the learner's mental structure (his presently available and favored way of thinking about things). Accommodation means adapting one's own mental structure by "taking cognitive account of the various real properties and relationships among properties that external objects and events possess." (Flavell, 1977; p. 7. Thus Piaget's assimilation-accommodation model provides a useful vehicle for thinking about how a person's cognitive system might gradually evolve with maturation and experience.

Four sequential stages of cognitive development are postulated in Piagetian theory; namely, sensorimotor, pre-operational, concrete-operational, and formal-operational stages. Each stage is characterized by an ability to perform certain mental functions and an inability to perform others (Kolodiy, 1975; Inhelder, 1974; Piaget, 1972).

Studies of the cognitive development of high school and college students, including the study analyzed for the present abstract, usually pertain to the concrete- and formal-operational stages which typically begin between ages 7-11 and 11-15 years, respectively.

When objects or events are used to provide a problem situation, the learner at the concrete-operational stage of cognitive development characteristically focuses on the perceptible and inferable reality in front of him. Various skills are used to order and interrelate whatever properties or features of the situation he or she can detect. His or her "conceptual approach generates solution attempts that are far more rational and task-relevant than those produced by the pre-operational learner. It does, however, pertain rather closely to detected empirical reality, and speculations about other possibilities occur only with difficulty and as a last resort" (Flavell, 1977, p. 103).

A learner at the formal-operational stage is more apt to begin with possibility and only subsequently proceed to reality. He or she may examine the problem situation carefully and try to determine what all of the possible solutions or states of affairs might be, and then systematically try to discover which of these is, in fact, the real one for the present case. For the formal-operational thinker, reality is seen as that particular portion of a much wider world of possibility which happens to exist or hold true in a given problem situation (Flavell, 1977).

> The formal-operational thinker inspects the problem data, hypothesizes that such and such a theory or explanation might be the correct one, deduces from it that so and so empirical phenomena ought logically to occur or not occur, and then tests his theory by seeing if these predicted phenomena do in fact occur.

> If you think you have just heard a description of textbook scientific reasoning, you are absolutely right. Because of its heavy trade in hypotheses and logical deductions from hypotheses, it is called hypothetico-deductive reasoning, and it contrasts sharply with the much more nontheoretical and nonspeculative empirico-inductive reasoning of concrete-operational thinkers (Flavell, 1977; pp. 103-104).

Research Design and Procedure

Two tasks, the solution of which required formal thinking, were presented to a total of 70 subjects who represented three distinct populations. One group consisted of 20 high school sophomores selected from an elective second-year BSCS biology course that was designed for students planning to pursue careers in science. A second group consisted of 25 college freshmen selected from introductory courses in physics and mathematics designed for science and engineering majors. The third group consisted of 25 college seniors majoring in the sciences.

SAT verbal and math scores were used as a basis for determining equivalency of the three groups. Since SAT scores were not available for the high school sophomores, scores of the previous year's graduating class were used to represent this group.

The test instrument consisted of two Piagetian tasks. One task titled "Combinations of Colored and Colorless Chemical Bodies" has been used by Piaget and other investigators (Inhelder, 1958) to measure ability to use combinatorial logic. The experimenter, in the presence of the subject, adds a few drops of a solution to a test tube containing a colorless mixture produced prior to the beginning of the experiment and notes the formation of a yellow color. The subject is then presented with five beakers of colorless liquids labeled g, 1, 2, 3, and 4. He or she is asked to reproduce the color and to find the function of each liquid.

A second task was entitled "Hauling Weight on an Inclined Plane" (Inhelder, 1958). The apparatus for the task consisted of a variable-height plane, a roller skate to which weights could be added or subtracted, a hanging weight, and a set of weights. The skate was set in equilibrium at the same angle on the plane. The subject was then asked to identify the variables that make up the equilibrium and to determine the relationship among them.

The percentages of students at each level (upper concrete, lower formal, upper formal) were summarized in tabular form. Differences between the three groups were analyzed via the nonparametric chi-square test with two degrees of freedom. This test provided the percentage of students at the formal level and below formal level. The expected frequencies were calculated from the criterion that the percentage of students scoring at the formal level in each group should equal the total percentage of all three groups scoring at this level. Also, a correlational analysis of all the variables was carried out.

Findings

1. The percentage of students at the upper formal level was roughly the same (35 and 32 percent) for high school sophomores and college freshmen, while somewhat higher (64 percent) for the college seniors.

2. Using the analog of Scheffe's theorem for chi-square, it was found that "the high school and college freshmen samples were equal and jointly different from the college senior sample" at the 0.05 level of significance.

3. A correlational analysis of variance revealed that scores on the Piagetian tasks were significantly correlated to SAT math scores but not to college grades which correlated significantly to SAT verbal scores.

4. A follow-up of the 25 college freshmen subjects showed that seven had failed to complete the second semester of physics. Each of the seven had scored as nonformal on the Piaget tasks.

Interpretations

The investigator concluded that a) a majority of high school and college students are below the formal level in cognitive functioning; b) college science education does not raise cognitive level but rather eliminates nonformal students through attrition; and c) college grades are more closely correlated to verbal ability than to cognitive functioning. He reasons that teachers should be aware of their students' level of cognitive functioning, and plan educational experiences accordingly. Also, that lecture methods must be balanced by more concrete activities where students can engage in manipulation of materials and verbal explanations among themselves.

ABSTRACTOR'S ANALYSIS

Since the primary purpose of this investigation was to "obtain indications of changes in mental level from high school through college," a longitudinal study that would have monitored the cognitive development of a given group of subjects over this four to eight-year period would have been optimal. Unfortunately, the time restrictions placed on most investigators, particularly doctoral candidates, make longitudinal studies impractical. Thus, a status study of *equivalent* groups of subjects is a reasonably valid practical alternative to a longitudinal study.

The present study is best classified as a "Static-Group Comparison" design according to Campbell and Stanley's criteria for designs for research (1963, p. 182). This pre-experimental design requires *equivalent* groups that are tested after having experienced different treatments. A major problem with the "Static Group Comparison" studies is that of determining that the groups are, in fact, equivalent.

SAT scores were used to establish group equivalency in the present study, a procedure that raises several questions relative to the actual equivalency of the groups. First, is this single criterian an adequate basis for determining equivalency when SAT scores were, in fact, not available for the high school sophomores? How valid is the use of SAT test scores of the previous year's graduating class (of the entire school?) to represent the scores of a group of 20 sophomores in an elective class for students planning college-preparatory careers in science? Are equivalent mean SAT scores indeed equivalent when the tests were taken over a period of four to six years during which national mean SAT scores declined annually? (Chronicle of Higher Education, Sept. 15, 1975: 18).

Second, were the subjects of the three groups from the same geographical region and from comparable socio-economic backgrounds? Had they attended comparable elementary and secondary schools? Piaget reports that the speed of cognitive development can vary from one environment to another, a fact that is corroborated by other researchers (Piaget, 1972).

Apparently Kolodiy assumed that the ages of the subjects is not a significant variable subsequent to the sophomore year (15 or 16 years). This may be questioned since Piaget has come to realize that the results he found with academically able adolescents from the better schools in Geneva who, at 15 and 16 years were demonstrating formal reasoning, cannot be generalized to other adolescent populations. He concludes that when it comes to formal thought, there may be a retardation in its formation to between 15 and 20 years when frequency and quality of intellectual stimulation received from adults is inadequate (Piaget, 1972; pp. 6-9).

An important omission of the journal article is a description of the basis for selecting the schools that the subjects were attending and the method of selecting the subjects from the universe of students in the classes from which the subjects were selected. Inferential statistics are based on the assumption of random selection of subjects from a population to which the findings are to be generalized. For the findings of the present study to be generalized validly to "the majority of high school and college students" (Kolodiy, 1975: p. 22), random selection of the subjects from this population would be essential.

The two tasks used as "The Test Instrument" have been used by Piaget and others (Inhelder, 1958) to a) assess ability to use combinatorial logic, and b) measure ability to identify the variables in an experiment: two aspects of formal reasoning. While no minimum number of tasks has apparently been established for a valid and reliable Piagetian test instrument, two tasks would appear to be a minimal number from which to infer three levels of cognitive development. Chiapetta's (1976) recent review of Piagetian studies relevant to science instruction at the secondary and college level reveals that three or more tasks are commonly used.

Although a general description of the tasks is presented in the article, the reader is not informed as to who conducted the interviews, how this person (these persons?) was trained, or details of the interview technique. Was a pilot study conducted to perfect the interview technique, to learn of the types of responses to expect, and to test the data-recording techniques? Were audio- or video-tapes used to record the interviews for subsequent analysis? In what order were the three groups of subjects tested and over what period of time? What criteria were used to classify the subjects as upper concrete, lower formal, or upper formal? While these questions may appear trivial, their answers would be essential to investigators desiring to replicate the present study.

The nonparametric chi-square test is appropriate for determining whether or not a difference existed among the groups and the correlation analysis if the variables studied provides valid information so long as no cause-and-effect relationship is inferred.

Kolodiy's conclusions are qualified in the text of the article by the phrase "within the limits of this study." Yet, he tends to generalize beyond the study. The sub-heading of the article generalizes to a "mismatch between the cognitive level of college students and the content and teaching of college courses." This may be misleading to the less than critical reader for the present study did not focus on the content and teaching of college courses. His conclusion that "the majority of high school and college students are below the formal level in cognitive functioning" generalizes to a population beyond that represented by the subjects of the present study.

The conclusion that "teachers should be aware of their students' level of functioning and plan educational experiences accordingly" is certainly consistent with accepted principles of educational psychology. Yet if, as Flavell (1977) argues, hypothetico-deductive reasoning and scientific reasoning are synonymous, should not advanced courses for science majors require formal reasoning abilities?

In summary, Kolodiy's study avoids a major criticism of many studies in education, the lack of a theoretical basis. His findings are generally consistent with and contribute to those of a rather large number of contemporary studies based on Piaget's theory of cognitive development at the formal-operational level (Chiapetta 1976). Unfortunately, as described by the journal article that serves as the basis for this analysis, the present study makes no new methodological contribution to Piagetian research. Nor does it provide a good model for other investigators because of the many omissions that have been cited which make replication problematic. Generalizing beyond the population represented by the subjects of the study is unfortunate in that these generalizations are often remembered and cited by the less-than-critical reader.

The difficulty in establishing equivalency of groups of subjects points to a need for longitudinal studies to determine what changes, if any, occur in mental level from high school through college.

REFERENCES

Campbell, Donald T. and Julian C. Stanley. "Experimental and Quasi-Experimental Designs for Research on Teaching," Ch. 5 in Handbook of Research on Teaching, N.L. Gage, Editor. Rand McNally and Company, 1963.

Chiapetta, Eugene L. "A Review of Piagetian Studies Relevant to Science Instruction at the Secondary and College Level." Science Education, 60(2): 253-261, 1976.

Chronicle of Higher Education. September 15, p. 18, 1975.

Flavell, John H. Cognitive Development. Prentice-Hall, Inc., 1977.

Inhelder, Barbel and Jean Piaget. The Growth of Logical Thinking from Childhood to Adolescence. Basic Books, Inc., 1958.

Kolodiy, George. "The Cognitive Development of High School and College Science Students." Journal of College Science Teaching, 5(1): 20-22, 1975.

Piaget, J. "Intellectual Evolution from Adolescence to Adulthood." Human Development, 15: 1-12, 1972.

Lawson, Anton E. and John W. Renner. "Relationships of Science Subject Matter and Developmental Levels of Learners." Journal of Research in Science Teaching, 12(4):347-358, 1975.

Descriptors--Developmental Tasks; Educational Research; *Intellectual Development; *Learning Theories; Science Education; Secondary Education; *Secondary School Science; *Scientific Concepts.

Expanded Abstract and Analysis Prepared Especially for I.S.E. by John T. Wilson, University of Iowa.

Purpose

The primary purpose was to assess the understanding of concrete and formal operational concepts by concrete and formal operational students in secondary school science classes. It was generally expected that a relationship would be found between students' scores on selected Piagetian tasks and written tests measuring concrete and formal operational science concepts. That is: the concrete operational students would be able to understand only concrete concepts while formal operational students would be able to understand both concrete and formal operational concepts.

Rationale

The study relied on conceptual differences between concrete operational concepts and formal operational concepts. In the study, concrete operational concepts were defined as "concepts whose meanings can be developed from firsthand experience with objects or events", while formal operational concepts was defined as "concepts whose meaning is derived through position within a postulatory-deductive system." Hence, the study assumes that a concrete concept may arise from direct sensory and tactile experience while formal operational concepts arise in a formally or logical deductive manner through imagination or through their logical relationships of ideas within a system.

As is common in studies of this type, emphasis was placed upon understanding formal operational concepts, rather than upon knowing them. Understanding in this study was limited to students being able to answer comprehension and application type questions based on Bloom's Taxonomy. Undoubtedly the study is drawn from other earlier studies in the general area of Piagetian research, however, the authors did not provide any clues as to which studies in this vast area of research were of particular importance. The basic reason for the study was to question the validity of contemporary high school science curricula that present formal operational type concepts to a population that is primarily concrete operational.

Research Design and Procedure

The design might be best described as an ex post-facto type design resembling more an after-the-fact assessment without any real treatment control. Measures were constructed and administered external to all classroom instruction. This included content tests as well as Piagetian tasks. The four Piagetian tasks included: conservation of weight, conservation of volume, separation of variables, and equilibrium in the balance. In scoring the Piagetian tasks numerical scores were awarded on a 1 to 5 scale with pre-operational being 1 point, transition to concrete 2 points, concrete 3 points, transition to formal 4 points and formal 5 points. Subject matter measures consisted of 15 concrete multiple choice questions and 15 formal multiple choice questions. Each set evaluated major concepts taught during the school year. Teachers did review the questions to check the representativeness of the subject matter included; this was identified as "content validity." A Spearman-Brown split correlation was used to estimate reliability of the subject matter tests. The Ss in this study were selected from one high school of over 2000 students and represent a generally above average, suburban population.

A number of statistical manipulations were used to assess the relationship between performance on Piagetian tasks and performance on tests. Included was a semipartial regression to assess which Piagetian task best predicted the Ss' scores on the concrete and formal operational portions of the written examinations. Although Ss were scored on a 1-5 basis on the Piagetian tasks, they were somehow assigned to one of seven levels of operational levels. It is not clear how this assignment was related to performance on the Piagetian tasks. Generally, data were summarized in terms of percentages of correctly answered concrete questions and percentages of correctly answered formal concept questions. In addition, a stepwise multiple regression table was included summarizing the data for the best predictor.

Findings

Of the biology sample, about two-thirds of the Ss were categorized as concrete operational or post-concrete operational. In the chemistry sample, most of the Ss were in transition between concrete operational and formal operational with almost all the sample categorized as post-concrete formal operational and transition to formal. In the physics class, Ss were generally between concrete operational and formal operational. However, the physics sample exhibited a flat distribution with Ss located in seven categories.

Concerning the relationships between students and their performance on the content area questions, most students in the lower levels of intellectual development found the tests particularly difficult while students in the higher levels found the tests easier as exhibited by the percentages of correctly answered questions. This pattern held true for concrete concept questions as well as formal concept questions. It was not at all surprising that all students found the formal concept questions more difficult than the concrete concept questions. Multiple correlation coefficients, which were not reported, were significant at the .05 and .01 levels. (One must assume that the authors mean significantly greater than zero. It is entirely possible that these may be significantly greater than zero but not account for much variance.) These unreported multiple correlation coefficients were summarized to suggest that Piagetian tasks are significantly correlated with understanding as operational as defined by the investigators and that there appeared to be more positive correlation between the tasks and their understanding of formal concepts and concrete concepts. The semipartial regression analysis disclosed that the conservation of volume task and the separation of variables task were the best predictors.

Interpretations

Generally the data were interpreted by the authors to support the premise that concrete operational Ss were unable to develop understanding of formal concepts. Also, concrete operational Ss were able to demonstrate understanding of concrete concepts. Formal operational subjects were considered to be able to develop understanding of both concrete and formal concepts. Additional interpretation of findings suggested that the apparent retarded development in the biology class may be traced to inappropriate subject matter and teaching procedures. In general, the authors suggest that a substantial portion of secondary school subject matter may not be suitable in terms of intellectual level of the learner, especially in the case of biology students who were unable to develop appropriate understanding of formal abstract concepts. It would appear that a science course dealing with abstraction and basic unifying themes is inappropriate.

ABSTRACTOR'S ANALYSIS

Little new information was provided by the study. It is not surprising that the students had difficulty answering difficult questions. Higher level questions (Bloom's Taxonomy) are more difficult for students to answer because they require more information for the student to think about in order to answer. It was not surprising that students on a concrete level find simpler questions easier to answer.

Many specific problems exist in the study: The statistical analysis, the definition of concrete-operational concepts, the method of assessing understanding, the lack of protocols for Piagetian tasks, and so forth. Attention should be given to research which attempts to link classroom achievement and instructional techniques to intellectual development. However, in constructing studies of this nature, comparisons should be made that will help teachers and curriculum developers modify the materials that they are using or producing. Certainly it is unreasonable to ask the curriculum developer to leave out all that "theoretical stuff of science." In many cases, the theoretical-stuff is the backbone of science. Science must be viewed as more than a collection of hands-on experiences: it must include the development of abilities to construct explanations. The concern of this type of research should not be to find out what concepts students are <u>not</u> understanding but rather to find out how students can be helped to learn those concepts.

Specifically, the following are some directions for research of issues in this study:

1. View carefully the methods of instruction in relationship to the learner's ability and level of development. In this study, no attention was given to the style and instruction.

2. Provide careful descriptions as to how data are managed and how <u>Ss</u> are assigned to groups. The researcher is obligated to explain his statistical, scoring, and other data management and collection activities.

3. Select prediction measures (tasks and other assessments) which fairly represent the operations or abilities necessary in learning the content. Then, remember that they relate to the <u>learning</u>, not always to the testing. What is known in this study is that some learners on concrete levels had a great deal of difficulty completing formal level test questions. Was the content also presented in a formal level? Were the learners unable to <u>learn</u> the concepts or only complete the test?

4. Treatment design and methodology need not be a problem. Much has been written in the area of Aptitude-Treatment Interactions (ATI). Many of the problems not solved in this study have been addressed in ATI literature.

5. Interpretation and recommendations must be reasonable in view of a study's limitation and realities of the classrooms, content areas, and teacher's skills.

Sayre, Steve; Ball, Daniel W. "Piagetian Cognitive Development and Achievement in Science." Journal of Research in Science Teaching, 12(2):165-174, April 1975.

Descriptors--*Academic Achievement; *Cognitive Development; *Developmental Tasks; Educational Research; *Learning Theories; Predictor Variables; Science Education; Secondary Education; *Secondary School Science

Expanded Abstract and Analysis Prepared Especially for I.S.E. by David J. Harmon, The Duval County School Board, Jacksonville, Florida.

Purpose

The stated purpose of the study was to determine the relationships between scholastic grades in science of junior and senior high students and the ability of students to perform formal operational tasks. There were six hypotheses tested.

Rationale

According to Piagetian theory, intellectual development evolves through four stages: (1) the Sensorimotor Stage (0-2 years), (2) the Preoperational Stage (2-7 years), (3) the Concrete Operational Stage (7-11 years), and (4) the Formal Operational Stage (11-16 years). It was assumed that the stages are of invariant sequence and may vary in duration and chronological age. Empirical support for purporting a relationship between cognitive development and achievement (science or in other disciplines) was not included. Elaboration of possible relationships between cognitive development and learning could provide implications for the science teacher and the teacher educator.

Research Design and Procedure

Four hundred nineteen science students (214, grades 7-9, and 205, grades 10-12) enrolled in eight Weld County (Colorado) Reorganized School Districts were randomly selected to participate in the study. Students successfully completing four or more of the five criterion tasks of the Piagetian Task Instrument (PTI) were classified as formal operational; others were classified as nonformal operational.

Four interviewers were trained to promote proper administration and evaluation of student performance on the PTI. Interviewer reliability (percent of agreement among interviewers) was established during a pilot study with 16 ninth grade students. The percent of agreement was 81 or higher for student performance on each task and overall. Face validity was obtained by using tasks modified from Inhelder and Piaget (1971), and Karplus and Peterson (1970). The reader should note the Inhelder and Piaget (1971) reference was cited but not included in the bibliography. The five tasks are listed, relevant equipment is identified and the interviewers' and students' roles are discussed below.

Task 1: Stickmen (Propositional Logic and Hypothetical Deductive Reasoning). The S was given an 8½ by 11 card with two stickmen drawn, one on each side of the card. The stickmen varied in height, one being two-thirds the height of the other. The S was asked to measure and record the height of the two stickmen with a set of eight connected "jumbo" paper clips. The "jumbo" paper clips were replaced with a smaller set of 12 paper clips. The S was then asked to measure only the shorter stickman and predict the height of the larger stickman.

Task 2: Pendulum (Combinational Logic and Hypothetical Deductive Reasoning). The S was given a simple pendulum consisting of a string, which could be shortened or lengthened, and a set of varying weights and asked to determine which variable(s) affect(s) the frequency of pendulum oscillation. The problem was to isolate string length from other variables.

Task 3: Balance (Propositional Logic and Hypothetical Deductive Reasoning). The S was presented a balanced scale. One weight was removed while the interviewer maintained equilibrium by holding the force arm. The S was asked, "Using any of the weights in front of you, how could you get the scale to balance?" After the S's response, another question was asked, "Are there other ways to balance the scale besides the one you chose?" The interviewer then removed the weight from the scale and placed another weight nearer the fulcrum while maintaining equilibrium by holding the force arm. The S was asked how he might balance the scale, and to justify his responses.

Task 4: Chemicals (Combinational Logic and Hypothetical Deductive Reasoning). The S was given four flasks containing perceptually identical liquids. The flasks (numbered 1-4) contained: dilute sulfuric acid, water, hydrogen peroxide, and thiosulfate. A fifth flask containing iodine was presented and labeled g. The S was given two glasses, one containing 1 + 3, the other containing 2. The contents were not revealed to the S. Several drops of g were poured into each of the two glasses. After observing the reactions, the S was asked, "How can you reproduce the color?" The S was allowed to attempt to reproduce the color. If successful, the S was asked to identify the functions of the liquids.

Task 5: Syllogisms (Propositional Logic and Hypothetical Deductive Reasoning). Three syllogisms were presented to the Ss, each consisting of two premises and a conclusion. After the S responded that the argument was valid or invalid, he was asked to justify his choice.

Scholastic grades were obtained for both formal and nonformal operational Ss at each of the six grade levels. Data were analyzed using statistical techniques, including Chi-Square Test of Independence and Point Biserial Correlation. A controversy as to whether scholastic grades are categorical or continuous data prompted the use of these two techniques.

Findings

Both formal level junior and senior high Ss received significantly higher ($p \leq 0.01$) grades than did nonformal Ss. In regard to tested hypotheses:

1. There was a significant relationship ($p \leq 0.01$) between scholastic success of eighth grade, ninth grade, biology and chemistry students and their performance on the PTI.

2. There was no significant relationship ($p \leq 0.05$) between scholastic success of physics and seventh grade science students and their performance on the PTI.

Significant correlations ($p \leq 0.01$) were found between the number of tasks performed at the formal operational level and science grades of junior high Ss ($r = 0.33$) and senior high Ss ($r = 0.46$). A significant relationship ($p \leq 0.05$) resulted between the scholastic science grades of both junior and senior high Ss and their performance on each of the tasks.

Interpretations

Students defined as formal operational received higher scholastic grades than did students defined as nonformal. Sayre and Ball state "...the findings seem to indicate a gradual growth of science students to complete formal operational tasks." They maintain that this study reaffirms Piaget's conceptions of the developmental growth of the intellect.

It was recommended that: (1) preservice teachers should develop greater understanding of Piagetian theory and (2) secondary science instruction should be structured around the cognitive developmental levels of the students involved.

ABSTRACTOR'S ANALYSIS

Pilot Study

The authors contend that interviewer reliability for the PTI was established during a pilot study. The pilot sample was limited to 16 ninth grade students enrolled in the University of Northern Colorado Laboratory School. The reader does not know: (1) the similarity of the ninth grade students to the main study sample (grades 7-12); or (2) if the percent of agreement among interviewers would be equally acceptable if grade levels (7-12) had been represented.

The Population and Samples

Two random samples of science students were selected, one at the junior high level (N = 214), the other at the senior high level (N = 205). Simple random sampling can only guarantee representativeness in terms of equality of probability for selection. This procedure does not guarantee inclusion of Ss from any particular subgroup of the population (e g., grade levels). A stratified random sampling technique would be needed to attain grade level representation. Several unanswered questions remain: (1) How was the sample size determined? (2) How many teachers and schools were represented by the selected Ss? (3) Were cultural differences that may affect cognitive development adequately controlled? A more elaborate pilot study would have provided implications for questions 1 and 3. Other comments related to sampling are mentioned in the next section.

Variables

The authors did not specify whether quarter, semester, or year science grades were employed. Nor did they provide an interview timetable. It is likely that internal validity was threatened by maturation.

Level of Cognitive Development. The method used to classify subjects may be open to question for several reasons:

1. Lunzer (1965) found that Ss who perform a task at one level of thinking may well perform the next task at a more or less sophisticated level, even if the two tasks are logically similar.

2. Beard (1962) found that for 8 to 16 year olds, the level of logical thinking varied extensively between schools. As with the study being reviewed, the ability to control variables was included.

3. Linn and Levine (1976) found that adolescents' (ages 11-16) success rates on solving logical problems depended on the familiarity of the variables.

4. No rationale was presented for using successful completion of four or five tasks as a cut-off point.

Perhaps developing a more specific scoring procedure and comparing student responses on logically similar tasks would have strengthened the authors' classification method.

Science Grades. This abstractor finds it difficult to think of a measure less reliable than scholastic grades, especially when collapsed within and over grade levels. A list of possible extraneous factors would be lengthy. Concerning grades, Schwebel (1972) noted that they are probably based on the ability to memorize, particularly if a nonformal student is confronted with formal subject matter.

Findings

The statistical techniques seem appropriate. Many of the findings were presented by school types; perhaps an hypothesis should have been stated for each (i.e., junior and senior high).

Finding c (page 171) regarding junior high science Ss is not supported by the data. A significant ($p < 0.05$) relationship was reported between the scholastic science grades of junior high students and their performance on each of the tasks. Table IV (page 170) indicates no significant relationship between scholastic grades of junior high students and their performance on either Task 4 or 5. Data in Table V (page 170) are less supportive.

With junior high students, finding a (page 170) indicates a correlation of .33 between number of tasks performed at the formal level and grades. Table IV (page 170) indicates that r = .38.

No significant relationship ($p < -0.5$) was obtained between grades of physics and seventh grade science Ss and performance on the PTI. In explanation, it was noted that 91.4 percent of seventh grade Ss were classified as nonformal and 80.7 percent of the physics Ss were considered formal operational. The strength of this explanation is reduced by the following: (1) Table VI (page 172) indicates that 90 percent of eighth and 86.5 percent of ninth grade Ss were classified as nonformal; and (2) 89.3 percent (abstractor's calculation) of all junior high Ss were nonformal. Perhaps a closer look at physics and seventh grade Ss is warranted.

Summary

This research attempts to bridge the gap between theory and practice. It is likely that the study has stimulated the interest of teachers in the possible relationships of individual differences and achievement.

REFERENCES

Beard, R. M. "Children's Reasoning." Mathematics Teaching, 21:33-39, 1962.

Karplus, R. and R. Peterson. "Intellectual Development Beyond Elementary School II: Ratio, A Survey." School Science and Mathematics, 70: 813, 1970.

Linn, M. C. and D. I. Levine. Adolescent Reasoning: The Development of the Ability to Control Variables. Advancing Education through Science Oriented Programs, Report PSc-9, Lawrence Hall of Science, Berkeley, 1976.

Lunzer, E. A. "Problems of Formal Reasoning in Test Situations." In P. H. Mussen (Ed.), European Research in Cognitive Development. SRCD Monographs, 30:19-46, 1965.

Schwebel, M. Logical Thinking in College Freshman. U.S. Department of Health, Education and Welfare, Office of Education, Bureau of Research, Final Project Report, 1972.

Lawson, Anton E., Anthony J. D. Blake and Floyd H. Nordland. "Training Effects and Generalization of the Ability to Control Variables in High School Biology Students." Science Education, 59(3):387-396, 1975.

Descriptors--Biology; *Critical Thinking; Educational Research; *Intellectual Development; *Learning Theories; Science Education; *Secondary Education; *Secondary School Science

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Thomas P. Evans, Oregon State University.

Purpose

The investigation was designed to study the following questions:

(1) Can the ability to control variables be taught to high school biology students who do not demonstrate formal reasoning on a written test of logical operations?

(2) Are students who are classified as early formal operational thinkers on the written test of logical operations able to benefit more from the training than students who are classified as early or late concrete operational thinkers?

(3) If the ability to control variables can be learned, is it generalizable to problems utilizing novel materials?

Rationale

The investigation was conducted using the Piagetian model of intellectual development as its theoretical framework. The investigation represented an attempt to test Piaget's position with regard to the acquisition of formal reasoning ability. It was related to earlier investigations by Case (1974), Bredderman (1972), Bass and Montague (1972) and Case and Fry (1973). The Piagetian tasks and scoring procedures used in the investigation were presented in detail in an earlier study by Lawson, Nordland and Devito (1974).

The major assumptions underlying the investigation were the acceptance of the position that the ability to control variables requires formal reasoning ability and the acceptance of the Longeot pencil and paper test of logical operations as a valid and reliable means for classifying students into selected Piagetian levels of intellectual development.

Research Design and Procedure

The researchers reported that a posttest-only design was used in the study. Sixty-five high school biology students whose ages ranged from 14 years, 7 months to 17 years, 10 months were randomly assigned to an experimental or a control group. The 33 students in the experimental group received training in the ability to control variables while the 32 students in the control group received no training.

Students in the experimental group were further subdivided into two groups prior to the training sessions. Students in each of the experimental subgroups worked in pairs and participated in four, 50-minute training sessions which followed the SCIS exploration-invention-discovery sequence and mode of instruction. Two problems involving the control of variables were used during the training sessions. The problems included the determination of the period of a pendulum and an investigation of variables related to the number of rotations of a rotoplane. During the first session, students participated in an exploration lesson using the pendulum. The concepts of variables and period of a pendulum were introduced in the second session. The concept of a controlled experiment using the pendulum was introduced in the third session. The fourth session was a discovery lesson. It provided the opportunity to apply the concepts of variables and controlled variables using a rotoplane.

The posttest consisted of interviewing each student in the control and experimental groups on three Piagetian tasks; i.e., exclusion of irrelevant variables, separation of variables and equilibrium in the balance.

Students in both the experimental and control groups were given the Longeot pencil and paper test of logical operations prior to the training sessions. The test reportedly allowed the classification of the students into concrete, transitional, and formal operational stages of development.

A t-test was used in analyzing differences between experimental and group mean posttest scores on the three Piagetian tasks.

Findings

The findings reported by the investigators were as follows:

1. Student responses in both experimental and control groups were classified on the Longeot test of logical operations as being 14 percent early concrete operational (IIA), 41 percent fully concrete operational (IIB), 35 percent early formal operational (IIIA), and 8 percent fully formal operational (IIIB);

2. The mean scores of the experimental group (15.2) and control group (14.9) on the Longeot test did not differ significantly at the 0.05 level;

3. The mean score of the experimental group (3.12) on the exclusion task was significantly larger at the 0.01 level than the mean score of the control group (2.53);

4. There was no significant difference between the mean scores of the experimental group (2.57) and control group (2.56) on the separation of variables task;

5. The mean score of the experimental group (2.57) did not differ significantly from the mean score of the control group (2.69) on the equilibrium in the balance task; and

6. Students in the experimental group who were classified IIA, IIB or IIIA on the Longeot test showed an overall average gain of 1.5, 0.69 and 0.33 substages, respectively, when classified on the Piagetian tasks following training.

Interpretations

The investigators concluded that training can increase student performance on a task designed to measure the ability to control variables when the task involves materials similar to those used in the training. The fact that the control group performed as well as the experimental group on the tasks measuring transfer of training suggested that the improved performance on the trained tasks was the result of rote learning rather than an increase in intellectual development.

Another stated conclusion was that students who were classified as early formal operational on the Longeot test did not benefit more from the training than students who were classified as concrete operational.

Upon analyzing why the training was not more successful, the investigators decided that the factor most likely missing from the training sessions was the self-regulation or equilibration factor. They proposed a hypothetical model of how self-regulation might operate in the training sessions and predicted that students would perform at an increased level if they participated in more experiences similar to those in the training sessions over a longer period of time and if they were encouraged to think through the problems at their own individual rates.

ABSTRACTOR'S ANALYSIS

Over the past 15 years, numerous researchers have examined whether or not training can result in the acceleration of students through the Piagetian levels of intellectual development. An analysis of these investigations have revealed that students can be taught to verbalize or perform specific tasks requiring higher developmental levels, especially when the verbalizations or tasks were similar to those introduced in the training sessions, but the ability to perform at higher levels was not generally retained or transferable to novel situations. A portion of the results of this investigation by Lawson, Blake and Nordland support the conclusions drawn from earlier investigations. It does, however, provide the added dimension of further supporting the notion that improved performance could be attributed to rote learning rather than an increase in intellectual development. Collectively, these investigations uphold the Piagetian prediction regarding acceleration of development of students from one intellectual stage to another.

A second finding reported by Lawson, Blake and Nordland did not uphold what the investigators stated as a Piagetian prediction. The prediction being that training would only be effective for students classified as transitional, defined in the study as late concrete operational and early formal operational. Students classified as early concrete and fully concrete operational showed greater improvement as a result of the training than did students classified as early formal operational. The investigators were quite surprised by this finding and suggested that it may have been the result of rote learning and ceiling effect. An additional rival hypothesis for explaining the finding and one not mentioned by the investigators may have resided in the quality of the pre- and post-test instruments and/or the interrelationship between what the instruments were purported to have measured. Unfortunately, a detailed description of the instruments, including discussions of their validity, reliability and intercorrelations were not included in the research report. The only bibliographical references made regarding the Longeot test were incorrectly listed and therefore unavailable.

The research report would have been improved if it had included additional information regarding the testing instruments. Some information on the tests was available, nevertheless, in a more recent report by Lawson and Blake (1976). They made comparisons among three tests, including the three Piagetian tasks, a biology examination, and a 19-item version of the Longeot test. Chi-square analysis yielded a significant relationship at the 0.02 level between student classification on each Piagetian level on the Longeot test and three Piagetian tasks. The results of the chi-square analysis presented by the investigators appeared impressive, but Light (1973) has pointed out that chi-square suffers from a lack of interpretability. It is possible to obtain a highly significant chi-square value even though the variables are only slightly associated. The chi-square value showed that a pattern existed between the test results that was significantly different from a random pattern, but it did not provide a measure of the correlation between student placement on each Piagetian level by the two instruments.

In fact, data from the contingency table revealed a skewed pattern with students being placed on the same Piagetian level on both instruments only 29 out of 60 times or roughly 48 percent of the time. Students were classified at higher Piagetian levels on the Piagetian tasks than on the Longeot test 23 out of 60 times or approximately 38 percent of the time. This apparent skewed relationship suggests a plausible explanation for the concrete operational students showing greater improvements than students classified as transitional or early formal operational as a result of the training sessions. Students classified as fully concrete operational on the Longeot test may have actually been transitional while those classified as being transitional on the Longeot test may have been fully formal operational as measured by the Piagetian tasks. Such an analysis reveals the necessity for an indepth examination of the testing instruments. It further suggests that the discussion by Light (1973) on the analysis of qualitative data in the Second Handbook of Research on Teaching be examined before accepting chi-square as a method for analyzing data in contingency tables.

One question should be raised regarding the chi-square analysis between student classification on the Longeot test and three Piagetian tasks beyond it being a questionable method of analysis. A recalculation by the reviewer revealed a chi-square value somewhat less than the reported value. The recalculated value was significant at the 0.11 level rather than the 0.02 level.

In addition to the investigation by Lawson and Blake (1976), the Longeot pencil and paper test of logical operations was described in a companion research report by Lawson (1975) and a more recent report by Lawson and Wollman (1976). The original test contained 28 items, but a shortened form was used in the previously identified reports. Lawson (1975) administered a 15-item version of the test, although he stated that he had reduced the test to 14 items. Lawson and Blake (1976) used a 19-item version of the test, and Lawson and Wollman (1976) shortened the test to 8 items. The use of the inconsistent number of items in these investigations and failure to identify which version of the test was used in the present investigation provides a very confusing situation. In addition, it is difficult to have confidence in a test when such procedures, coupled with expressing the reliability in terms of internal consistency and not re-establishing validity, are used. Modifying can and frequently does have serious effects on test reliability and validity.

An explanation of why the research design was referred to as a posttest only design also would have improved the research report. Experimental and control group posttests were used for answering the first question, but the Longeot test apparently served as a pretest and three Piagetian tasks were used as a posttest for answering the second question of which level of students benefitted most from the training. The Longeot test was not called a pretest; however, it was used to classify students into the same Piagetian levels as did the three Piagetian tasks.

The recognition and statement by the investigators that the appropriateness and/or quality of the training sessions could have been the reason for the lack of success or the posttests by the experimental group was commendable. The writers of other research reports would do well to follow this example when it is applicable.

The prediction that the investigators generated from their hypothetical model of how self-regulation might operate in the training sessions was interesting and congruent with the Piagetian model of intellectual development. Of course, it remains as a proposition requiring further research. Additional areas of research include determining the test-retest reliability and establishing the validity of modified versions of the Longeot test, comparing various methods of analysis of data in contingency tables, determining whether or not understanding of the concept of controlling variables is characteristic of high school biology students and examining different strategies for training students to control variables.

REFERENCES

Bass, J. E. and E. J. Montague. "Piaget-Based Sequences of Instruction in Science." Science Education, 56:503-512, 1972.

Bredderman, T. A. "The Effects of Training on the Development of the Ability to Control Variables." Journal of Research in Science Teaching, 56:503-512, 1972.

Case, R. "Structures and Strictures: Some Functional Limitations on the Course of Cognitive Growth." Cognitive Psychology, 6;544-573, 1974.

Case, R. and C. Fry. "Evaluation of an Attempt to Teach Scientific Inquiry and Criticism in a Working Class High School." Journal of Research in Science Teaching, 10:135-142, 1973.

Lawson, Anton E. "Sex Difference in Concrete and Formal Reasoning Ability as Measured by Manipulative Tasks and Written Tasks." Science Education, 59:397-405, 1975.

Lawson, Anton E., Floyd H. Nordland and Alfred Devito. "Piagetian Formal Operational Tasks: A Crossover Study of Learning Effect and Reliability." Science Education, 58:267-276, 1974.

Lawson, Anton E. and Anthony J. D. Blake. "Concrete and Formal Thinking Abilities in High School Biology Students as Measured by Three Separate Instruments." Journal of Research in Science Teaching, 13: 227-235, 1976.

Lawson, Anton E. and Warren T. Wollman. "Encouraging the Transition From Concrete to Formal Cognitive Functioning-An Experiment." Journal of Research in Science Teaching, 13:413-430, 1976.

Light, Richard J. "Issues in the Analysis of Qualitative Data." In Travers, Robert M. W. (Ed.) Second Handbook of Research on Teaching. Chicago: Rand McNally College Publishing Co., 1973.

Espejo, Mila, Ronald Good, and Paul Westmeyer. "Evaluation of a Child-Centered Science Curriculum Using the Intellectual Models of Piaget and Guilford." Journal of Research in Science Teaching, 12(2):147-155, 1975.

Descriptors--*Curriculum; Educational Research; Elementary Education; *Elementary School Science; *Learning Theories; Science Education

Expanded Abstract and Analysis Prepared Especially for I.S.E. by William R. Brown, Old Dominion University.

Purpose

The purpose of this investigation was to determine the effectiveness of a child-structured science program in further developing selected intellectual factors that appear to be sequentially developed among pre-operational and concrete-operational children. A major component of the study was the development of tasks to measure the hypothesized intellectual factors.

Rationale

The investigators proposed that instructional programs should be evaluated in order to determine the effectiveness of the materials and procedures in attaining stated goals. Research findings suggest that Guilford's structure-of-intellect model presents the possibility of isolating specific intellectual skills that can be further developed by specific learning experiences. The experiences should be geared to the cognitive capacity and mode of functioning of children as described by Piaget.

Research Design and Procedure

The variate was student experiences in a child-structured science program. Child-structured was operationally defined as having the following characteristics: (a) Manipulative materials were available to every student during each lesson in quantities such that sharing was not required; (b) Sets of materials allowed for a variety of activities which were determined by the students; (c) The teacher avoided evaluation and directive behaviors and attempts to interact with individuals.

Intellectual factors hypothesized to be strongly associated with selected sets of science materials were: cognition, convergent production, and evaluation of figural classes; cognition and evaluation of figural relations; cognition and convergent production of figural implications; and evaluation of figural systems.

The criterion variable was performance on tasks designed for the study that measured the eight hypothesized intellectual factors. Factors in the test battery had a high probability of being developed as suggested by the nature of the activities. There had to be as many task variables in the test battery as were needed to identify and qualify the factors in the factor matrix. The tasks and the respective factors are discussed in detail in the report.

Two tasks were used only to indicate whether a child was at the pre-operational stage in the development of class concept. A third task was used to determine whether a child was at the pre-operational or concrete-operational stage of class concept development.

For size, area, and volume relationships, three tasks were designed to designate a child as being in the pre-operational stage. A fourth task assessed whether a child had passed into the concrete-operational stage. Another task involved the evaluation of serial systems and could not be used for stage categorization. The final task was used to indicate whether the child was in the concrete-operational stage of the development of the concepts of area and volume, respectively, or in the stage preceding it.

A reliability estimate of 0.927 was obtained for the cognitive test instrument by the use of alternate components.

A delta index was developed for use in validation. If the rate of development of the intellectual factors varies, scores of children within a particular age group on items homogeneously measuring the same factor would be expected to cluster on a development continuum. There would be as many clusters as factors being measured.

Task analysis was conducted to identify the tasks that contributed notably to the significant differences among the total test means of the subjects in each of the four treatment-age level combinations by analyzing for effects that could be attributed to treatment, age, or treatment-age interaction.

The treatment group was 23 first grade students. Two control groups consisted of 23 kindergarten students and 27 first graders. Intact classrooms were used. Age and socioeconomic status were used to screen students to be included in the project. The design model was treatment-posttest for the one class and posttest only for the control groups.

___ X 0

___ ___ 0

Toward the end of the school year students were interviewed on the tasks and responses were tape-recorded and scored.

Where the means on a given task of at least two of the four age-treatment groups were below 0.90, a two-way ANOVA was performed. t-Tests were used to compare means of all possible pairs of treatment-age groups to determine sources of significant differences. One-way ANOVA was used to investigate age effect for those tasks where means of three of the four treatment-age groups were above 0.90.

Findings

All the children were observed to be in either the pre-operational or the concrete-operational stage. There was a significant difference (.01) between the treatment and the combined control groups, but age and treatment-age interaction was not significant. Results favored the treatment group in terms of the development of the concept of class, and size, area, and volume relationship as well as the intellectual factors related to these concepts.

Interpretations

The results support the conclusion that an activity-centered program, where the experiences were designed to match the cognitive structures of children in certain stages of development, promotes the development of intellectual factors which enable children to move to the concrete-operational stage.

Results of the item analysis may be interpreted to imply that the activities on classification and size relationships promote the development of the intellectual factors: convergent production of figural classes and evaluation of figural systems, respectively. Cognition of figural classes and evaluation of figural relations may be developed by the time children are about seven years of age without specific experiences such as those provided in a child-structured program.

Intellectual factors from the Guilford model can be predicted, identified, and assessed by using interview tasks and techniques derived from the work of Piaget.

ABSTRACTOR'S ANALYSIS

This study dealt with the application of learning and cognitive development theories to classroom instruction. This has been indicated as a top priority for research by NARST members (Butts, 1977, p. 163). The integration of the Guilford and Piaget models is an excellent "start" at application of learning theories to science program planning.

The investigators developed and validated a series of tasks linked to specified intellectual factors. The description of the tasks in the report would allow for replication of the study. The replicability potential of the study is an asset not found in many educational studies.

Prior to replication or extension of the study, two points need to be clarified in the research design. First the length of time of the treatment should be specified. The investigators developed an excellent operational definition for a child-structured science program, but the reader is not informed as to the duration of the treatment. Treatment length and related variables such as specific teaching strategies used could be variables that affect performance on tasks.

A second design restriction is the lack of random assignment of subjects to treatment and control situations. Randomization could certainly strengthen confidence in the results.

It would be helpful if the "delta index" referred to for use in validation were explained in a summary fashion. A footnote specified that this information may be obtained from the investigators. The conclusions reached are dependent, in part, on a valid assessment instrument.

Unfortunately, many "advocates" of Piagetian principles are enamored with age guidelines. The investigators in this study were careful in this respect and established two levels with 6:11 as the dividing line. Reference to stages rather than ages lends credibility to the findings.

A fundamental question of this study is <u>what actually was evaluated</u>? The investigators refer to "evaluation of a child-structured science curriculum." Their working definition of "curriculum" actually describes instructional variables. If curriculum is "a plan, a structured series of intended student outcomes or a set of planned experiences indicating that which has been identified as worthwhile for students to learn or experience," then curriculum evaluation is accomplished through the consistent and logical use of statements in decision making (White, 1977, pp. 3-6). The assessment of student learnings is a procedure for evaluating instruction. Instruction refers to the behavior of teachers and the strategies selected to promote the objectives of the curriculum. It appears to this reviewer that the intellectual factors as identified by Guilford can be used in a curriculum plan. How "well" these factors are developed depends on instruction such as defined in the tasks developed for this study.

The evidence gathered by this investigation supports the utility of the intellectual models of Guilford and Piaget in program evaluation for young children. Development of tasks for older learners would certainly be a useful contribution to the research pool.

REFERENCES

Butts, David, et al. "Priorities for Research in Science Education." Paper presented at NARST-1977. Abstracts of Presented Papers, NARST-1977, NARST, ERIC-SMEAC, Columbus, Ohio, 1977.

Goodlad, John. The Development of a Conceptual System for Dealing with Problems of Curriculum and Instruction. Los Angeles: University of California, 1966.
ED 010 064.

MacDonald, James. "Structures in Curriculum." In Theories of Instruction. Washington, D.C.: Association for Supervision and Curriculum Development, 1965.

White, Edwin, William Brown, et al. Handbook for Principals and Supervisors-Decision Making for the Elementary Science Program. Unpublished draft, Department of Education, Commonwealth of Virginia, Richmond, 1977.

Raven, Ronald J., and Robert Guerin. "Quasi-Simplex Analysis of Piaget's Operative Structures and Stages." Science Education, 59(2):273-281, 1975.

Descriptors--*Cognitive Development; Cognitive Processes; *Discovery Learning; Elementary Education; *Elementary School Science; Educational Research; Learning; *Learning Theories; Science Education; Socioeconomic Influences

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Anton E. Lawson, Arizona State University.

Purpose

The purpose of this study was to analyze the hierarchical scheme of seven operative structures defined by Jean Piaget as well as determine if the sequence of two substages of concrete thought and two substages of formal thought constitute a hierarchy of related components. Three hypotheses were tested:

1. Is there a hierarchical structure among the concrete operational stages III A and B and the formal operational stages IV A and B?

2. Is there a hierarchical structure among the logical operations of classification, seriation, logical multiplication, compensation, proportions, probability, and correlations?

3. Is there a factor describing stage III A operative structures and another factor describing stage III B, IV A, and IV B operative structures?

Rationale

The operative structures defined by Piaget are important because investigations have shown that they determine the form and function of concept acquisition. As an individual acquires more complex operative structures, he is capable of acquiring more complex concepts.

Piaget's theory of development states that the development of more complex operative structures depends upon prior development of simpler operative structures. If the hypothesized hierarchy of operative structures of classification, seriation, logical multiplication, compensation, proportions, and correlations is confirmed then it means that the K-12 science curricula should consider this sequence in designing instructional sequences for science concepts. Conversely, if the sequence is not confirmed then science instruction need not consider the interlocking dependence of these structures.

Research Design and Procedure

Instrument—the Raven Test of Logical Operations contains 68 pencil-paper items testing for use of the seven operative structures investigated. The test is divided into three parts each requiring about 45 minutes for administration.

Subjects were 896 male and female students ranging in age from 8.8 years to 19.4 years. Subjects were selected from third, fourth, fifth, sixth, seventh, and ninth grades and college freshman classes drawn from four metropolitan areas: Philadelphia, Pennsylvania; Buffalo, New York; Salamanca, New York; and Atlanta, Georgia.

Guttman's radex theory as modified by Schoeman was used as a quantitative technique to test the first two hypotheses. The third hypothesis was tested by use of Alpha Factor Analysis.

Findings

Hypothesis one was accepted in that the predicted order of complexity (stage III A, III B, IV A, IV B) was found. The rank order of the simplex loading coefficients associated with the substage levels was the same as the predicted order.

Hypothesis two was rejected in that the rank order of the coefficients associated with the seven logical operation variables was not the same as the predicted order.

Hypothesis three was accepted in that the expected loading pattern of the seven logical operations variables was observed. Classification, seriation, and compensation showed moderate loadings on factor II while logical multiplications, compensations, proportions, probability, and correlation showed moderate loadings on factor 1.

Interpretations.

The results of the study show that two concrete substages and the two formal substages of cognitive development are hierarchically related but that some overlap exists. Presumably some operations develop in a parallel fashion. The seven logical operations do not by themselves form a linear hierarchy. This result suggests that several logical operations within a group may be used to construct science concepts by individuals at a specific stage. The operative structures may be viewed as a means for extending the use of any given science concept within a given level of development thus enhancing the concept's generality and retention.

ABSTRACTOR'S ANALYSIS

This investigation was based primarily upon the rather tenuous hypothesis that the seven operative structures of classification, seriation, logical multiplication, compensation, proportion, probability, and correlation form a sequential hierarchy in the order listed. Raven and Guerin admit that Piaget himself does not suggest this hierarchy, nevertheless they claim that it is implied by Piagetian studies.

Certainly some sequentiality must exist (e.g., classification and seriation must come before correlation) but I find no justification for drawing such an implication from the studies cited by the present authors. A more modest but much more believable interpretation of those studies is simply that the logical operations of the concrete stage must precede those of the formal stage. This in no way implies that any two operations of a single stage, such as proportions and probability, are sequentially acquired within a given stage.

Raven and Guerin proceed from this tenuous hypothesis to state that if the hypothesis is confirmed, then the K-12 science curriculum should consider the sequence in teaching concepts. Further they state that if the order is not confirmed, then science instruction need not consider the interlocking dependence of these structures. For example, no attention need be paid to compensations and seriations when teaching concepts using proportions. I find this statement unwarranted and unjustified. To imply that the usefulness of Piaget's theory of the development of operative structures rests upon the confirmation of the present hypothesis of sequentiality is presumptuous.

Nevertheless, the idea of a sequence of hierarchically related operative structures within stages is an interesting one. Unfortunately it is not one so easily tested. The use of the factor analytic techniques employed by Raven and Guerin, although useful, cannot really tell us if, say, classification precedes or follows seriation or if proportions precede or follow probability. This is because the analyses are dependent in part upon the difficulty level of the items. The most difficult items generally turn out to be the ones at the top of the hierarchy. But this need not imply that they develop last. This is simply because not all items involving one operative structure are of equal difficulty. For instance, a number of items involving proportions and probability can be selected with a wide range of overlapping difficulty. If one selects only the easy proportions items and only the difficult probability items, then the statistical analysis will show probability at the top of the hierarchy. If, on the other hand, only easy probability items and difficult proportions items are selected then the order will be reversed. So which comes first? The analysis cannot tell us. For this reason hypotheses such as this are extremely difficult, if not impossible, to test adequately.

Raven and Guerin claim to have found support for their first hypothesis concerning the hierarchical structure of the concrete III A and III B and the formal IV A and IV B substages. What they in fact found support for was the following sequence: classification and seriation (which they placed in substage III A), logical multiplication and compensation (which they placed in substage III B), proportions and probability (placed in substage IV A), and correlations (placed in substage IV B). They have simply grouped test scores from different parts of the test. Empirical support for the validity of the sequentiality of this grouping of items is found but at this point a question is raised as to the value of such empirical support.

Let me clarify with an example. Take correlations, for instance. If one analyzes what is needed to solve a correlation task, it becomes immediately obvious that the data under question must be classified, then the number of instances of class members be compared, then joint probabilities or ratios of occurrences and nonoccurrences of class members must be computed. Without these initial steps the correlation task cannot be solved. Obviously then someone would not solve a correlation task if he were unable to solve the logically less complex task of classification and so on. This simply stands to reason. The empirical result that the correlation tasks are more difficult and are solved only by those who can also solve classification tasks is not the least bit surprising. Other similar arguments could be made for the other operative structures investigated.

The results of the test of hypothesis three that showed the operative structures do contain two general factors support earlier factor analytic studies which indicate two general modes of thought: concrete and formal. Raven and Guerin's distinction between the concrete factor as one involving the arrangement of events and objects by property and the formal factor as one concerning the coordination of changing variables appears insightful and potentially helpful. However, the means for using these two operative structure domains for the improvement of teaching science concepts suggested by the authors is not at all made clear.

Lawson, Anton E., Floyd H. Nordland and Alfred Devito. "Relationship of Formal Reasoning to Achievement, Aptitudes, and Attitudes in Pre-service Teachers." Journal of Research in Science Teaching, 12(4): 423-431, 1975.

Descriptors--*Achievement; *College Science: Educational Research; Higher Education; *Intellectual Development; Learning; Preservice Education; *Science Education

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Joseph P. Riley, II, University of Georgia.

Purpose

1. To determine relationships among students' scores on four Piagetian styled tasks measuring formal reasoning ability and scores on measures of attitude, aptitude, achievement and knowledge of the science processes.

2. Analysis of these interrelationships to determine if the Piagetian-styled tasks do reflect a meaningful measure of the extent of formal operational thinking abilities or if they simply measure physics or science content.

Rationale

Recent studies indicate that as many as 50-75 percent of secondary and college level students fail to demonstrate formal reasoning ability as measured by Piagetian-styled tasks. The authors raise a question about the validity of these findings, suggesting the possibility that the tasks may be dominated by physics content which is simply unfamiliar to students.

Research Design and Procedure

An *ex post facto* design using hypotheses testing procedures was employed. Four Piagetian-styled tasks were used to measure the formal reasoning abilities of 71 college freshmen and sophomore elementary education majors. The tasks and their reported reliability coefficient are: The Conservation of Volume Using Clay (.24), The Conservation of Volume Using Metal

Cylinders (.48), The Separation of Variables (.79) and The Exclusion of Irrelevant Variables (.79).

Points were assigned to each task and subjects stratified into three categories based on the sum of their total scores. These procedures identified 13 subjects as concrete operational, 47 as transitional, and 11 as formal operational.

The following achievement, aptitude, and attitude measures were obtained from student records and used as concomitant variables.

A. Achievement measures

1. Science - Biology/Chemistry College Entrance Examination
2. Science - Sequential Test of Educational Progress
3. Science - The Wisconsin Inventory of Science
4. Mathematics - College Entrance Examination
5. English - College Entrance Examination
6. High School Graduating Class Rank
7. College Gradepoint Average

B. Aptitude measures

8. Mathematics - Scholastic Aptitude Test
9. Verbal Scholastic Aptitude Test

C. Attitude measures

10. Attitudes toward Science and Science Teaching - The Bratt Attitude Test

The data were analyzed in three ways:

1. Intercorrelations were run on the task scores and the total task score with all achievement, aptitude, and attitude measures. Correlations which reached significance at the .05 level were corrected for attentuation.

2. One-way analysis of variance procedures were used to test for significant differences among the means of the concrete, transitional, and formal operational groups on each of the dependent variables.

3. Principal Components Analysis was run to determine the number of significant factors among the variables measured. Only loadings greater than .30 were included.

Findings

The majority, 66 percent of the sample, demonstrated transitional responses. With the exception of the conservation of Volume Using Clay task, results obtained on the Piagetian-styled interviews were consistent with previous studies. On the Clay task, 90 percent of the students exhibited conservation reasoning compared to previously reported percentages of 58 by Elkind (1962) and 61 by Towler and Wheatly (1971). Except for this same task, all other task scores correlated positively with all of the other measures. The correlations among students' total task score and all other measures were found positive and significant.

Correlations among the task scores and the science achievement tests as well as the mathematics and English achievement tests were significant, with the science measures somewhat higher. The correlations of task scores with the verbal and math aptitude scores were moderate to high. Moderate correlations were found between the Piagetian scores and the attitude test. Low to moderate correlations were found between the task scores and scores of the test of science processes.

Results of the one-way analysis of variance showed differences between the means of the three groups, with the formal group scoring higher than the transitional group and the transitional higher than the concrete group. However, only four of these differences were determined to be significant. The F-ratios on the Science (STEP) test ($p < .001$), the verbal aptitude test ($p < .01$), the math aptitude test ($p < .05$), and the attitude measure ($p < .05$) were found to be significant.

Factor analysis identified four components, accounting for 63.3 percent of the total variance. The first component, accounting for 33.1 percent of the variance, was identified as an achievement dimension. The Conservation of Volume Using Cylinders task was the only task which loaded on this component. According to the authors, this suggests that obtaining good grades is probably dependent upon something other than the logical reasoning abilities measured by these tasks. The second component was identified as a logical reasoning and attitude dimension (13.1 percent). The Separation and Exclusion tasks loaded significantly on this component. The authors suggest that it is possible that as a person develops the abilities to successfully respond to the Piagetian tasks his attitude toward science and science teaching as measured by the attitude instrument also improves. A science and mathematics achievement dimension was identified as the third component. The highest loadings were on the science achievement tests with .85 and .79. Math achievement and aptitude loaded at .53 and .45 respectively while English achievement and verbal aptitude loaded at .37 and .45 respectively.

Interpretations

The authors conclude that Piagetian measures of formal operational reasoning abilities are significantly related to achievement, aptitude, attitude, and knowledge of the science processes. They also conclude that the Piagetian tasks are generally, but not entirely, content free. The authors interpret the implication of their findings by suggesting that those secondary school and college students limited to concrete modes of thinking should first be confronted with ideas and materials at their own level of capability and then gradually be asked to deal with ideas at a more abstract or formal level.

ABSTRACTOR'S ANALYSIS

This *ex post facto* study supports previous research findings indicating little or no relationship between science knowledge and formal operational tasks. The results also contribute to existing research by providing further evidence of the construct validity of selected Piagetian styled, formal operational tasks. The strength of these findings have been evaluated in terms of the study's procedural and external validity.

Procedural Validity

Through hypothesis testing, the authors have avoided the unsystematic search for relationships so common to most *ex post facto* research. Instead, they have designed a controlled inquiry by predicting the significant and non-significant relationships they would have expected if their hypothesis were tenable. Procedures such as reporting the reliability of the Piagetian tasks, and limiting correction for attenuation to only those correlations reaching the .05 significance level, enhance the study's procedural validity. The analysis is appropriate and well reported but is not as complete as it could be. Post hoc analyses of the significant F ratios found in the analysis of variance are omitted. The authors' reported significant differences among the overall mean scores of the concrete, transitional, and formal groups on four of the measures. They also report that, in all cases, the mean scores of the concrete group were lower than those of the transitional group and the mean scores of the transitional group were lower than those of the formal group. The significant F ratios indicate that differences exist among these three sets of mean scores. They do not identify where these differences occur. Multiple comparison techniques should have been employed to identify where the significant differences could be found. The fact that the three sets of mean scores fell in the order they did, provides evidence that the differences are of the nature hypothesized. However, multiple comparison procedures could have provided more information on the strength of this evidence. For example, below are three possible results of multiple comparison analysis of these data:

A) $\bar{X}$ Formal $\bar{X}$ Transitional $\bar{X}$ Concrete

B) $\bar{X}$ Formal $\bar{X}$ Transitional $\bar{X}$ Concrete

C) $\bar{X}$ Formal $\bar{X}$ Transitional $\bar{X}$ Concrete

(means not underlined by the same line differ significantly)

In this example, C provides the strongest evidence that the results are as called for in the hypotheses. The results of B provide less evidence than C but more than A.

Despite the omission of multiple comparison procedures the results of the three different analyses; correlation, one-way analysis of variance and principal component analysis combine to provide strong arguments and logical validity for the study's conclusions.

External Validity

The heterogeneity of tasks, testing procedures, scoring and sample populations in formal operational research has made generalizability of findings extremely tenuous. Possible sources of threats to the external validity of this study are: the selection and scoring of tasks, and the sample population.

Tasks. The conservation of volume using metal cylinders task has been criticized for not testing the concept of conservation of displacement volume from a Piagetian standpoint (1977). Surprisingly, the results of this study provide more support for the use of this particular task than for one of the more traditional tasks of displacement using clay.

The authors indicate they used testing procedures elaborated by Piaget and Inhelder (1971) and Karplus and Lavatelli (1969). These two groups of researchers use different criteria for task responses. Piaget's testing procedures make use of an objective response as well as rigorous questioning of the subject to determine his justification. The latter group generally only require an objective response or an objective response with a brief comment of justification. The differences between these procedures can have a profound effect on conclusions (1975).

Sample. The authors state that they have no reason to suspect that in other samples of college or high school students similar positive correlations would not be found. However, lack of adequate information about the subjects raises questions as to their representativeness. The description of the sample is limited to the subjects' mean age and the fact that they were freshman and sophomore elementary education majors. No information is given about previous experiences of the subjects or how they were selected. Is it safe to assume that this highly select population of education majors had no previous contact with Piagetian theory? Previous experience with the more common Piagetian tasks could explain their highly skewed response on the conservation of volume using clay task. Information reconstructed from the tables raises other questions about the general representativeness of the subjects. The sample size is 71. However, the degrees of freedom on the analysis of variance table indicates 38 scores on the science achievement test. A note after the description of the science test states six students elected the biology achievement examination, the remaining elected the chemistry examination. How representative is a sample of 38 subjects of whom, when given a choice between taking a biology or chemistry test, 32 elect the chemistry test?

While some threats to external validity of formal operational studies can be eliminated by detailed research reporting, others resulting from the early state of the art in this area are unavoidable. Until there is more standardization in formal operational research, generalizability of results will be extremely limited.

Through *ex post facto* research methodology this study identifies the existence of relationships between formal operational reasoning and achievement, aptitudes and attitudes. In addition to contributing to the construct validity of formal operational development, it identifies relevant areas and variables worthy of further study. The search for relationships in formal operational studies, characteristic of the early stages in any research effort, should now give way to experimental investigations. The structural components of formal operations, how they develop and the variables which affect them are identified directions for further experimental research.

REFERENCES

Elkind, D. "Quantity Conceptions in College Students." Journal of Social Psychology, 57:459-465, 1962.

Karplus, R. and C. Lavatelli. The Developmental Theory of Piaget: Conservation, 1969. San Francisco, California: John Davidson Film Producers (1[illegible]57 Union Street).

Neimark, E. "Intellectual Development Through Adolescence." In F. Hirowitz (Ed.), Review of Child Development Research. Chicago: University of Chicago Press, 1975.

Phillips, D. "The Karplus Islands Puzzle, Volume with Metal Cylinders, and Other Problems." Journal of Research in Science Teaching, 14(4):361-369, 1977.

Piaget, J. and B. Inhelder. "Le Developpement des Quantites Physiques Chez L'enfant." Neuchatel: Delachaux and Niestle, 118:265-270, 1971.

Towler, J. O. and G. Wheatly. "Conservation Concepts in College Students, A Replication and Critique." The Journal of Genetic Psychology, 118:265-270, 1971.

INSERVICE EDUCATION

Hasan, Omar E., and Victor Y. Billeh. "Relationship Between Teachers' Change in Attitudes Toward Science and Some Professional Variables." Journal of Research in Science Teaching, 12(3):247-253, 1975.

Descriptors--Attitudes; Educational Research; *Measurement; *Science Education; *Secondary School Teachers; *Scientific Attitudes; Summer Workshops; *Teacher Background

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Frank A. Smith, West Chester State College.

Purpose

This investigation was an attempt to determine if a four-week summer training course in science teaching had an effect on science teachers' attitudes toward science and also whether certain teacher variables had an effect on these attitudes. The following null hypotheses were investigated:

Hypothesis 1: There is no change in teachers' attitudes toward science as a result of a four-week summer training course in science teaching.

Hypothesis 2: There is no significant association between change in teachers' attitudes toward science, as a result of a four-week summer training course in science teaching, and the variables: number of years of science teaching experience, number of years of college education, and number of weeks of relevant professional inservice training.

Rationale

The development of favorable attitudes toward science is generally agreed upon to be an important characteristic of a scientifically literate person. The authors adopt the assumption that students' attitudes toward science are related to their teachers' attitudes. Therefore, it is desirable to know what variables are related to favorable teacher attitudes.

Research Design and Procedure

The sample used in the study consisted of 129 secondary science teachers in Jordan who had been invited to attend a four-week training course in science teaching. The training course consisted of lectures and demonstrations in scientific concepts, teaching methods, and the nature of science, laboratory investigations, reading, and related films. The dependent variable, attitudes toward science, was measured by an attitude scale written by one of the authors. This instrument consisted of 32 items which were previously rated on a scale of 1-11 by 46 university science professors. The split-half reliability of the instrument was measured to be .62. The independent variables of the study were the training program, the number of years of science teaching experience, the number of years of college education, and the number of weeks of relevant professional inservice training. The research design was of the one-group pretest-posttest type. Mean scores on the attitude scale were calculated. The effect of the training program was determined by comparing pretest and posttest scores by means of a t-test. The effect of the various teacher variables was determined by multiple regression techniques.

The investigators found that there was no significant difference between pretest and posttest scores on the attitude scale as a result of the training program. The analysis did show, however, that two variables: previous professional inservice training and level of education were significantly ($p < .01$) related to teachers' gain scores on the attitude scale.

Findings

On the basis of these findings the investigators concluded that:

1. the four-week training program was ineffective in producing a significant positive change in teachers' attitudes.
2. teaching experience was not significantly related to teachers' changes in attitudes.
3. there was a positive relationship between previous inservice training and the teachers' change in attitude.

4. there was a negative relationship between level of education and teachers' change in attitude.

Interpretations

In a discussion of these conclusions the investigators suggested that the positive relationship between professional inservice training and teachers' change in attitudes could be explained as follows:

a) the professional inservice training to which these teachers had been exposed included lectures on the nature of science and the viewing of scientific films. Teachers with more exposure to such training might be expected to show more positive attitudes toward science.

b) the sample was largely volunteer and perhaps teachers who volunteer for such programs tend to be more susceptible to positive changes in attitude.

The negative relationship between level of education and teachers' change in attitudes suggested to the investigators that those with higher levels of education might have formed a stable structure of attitudes towards science which would be hard to change, while those of lower levels might not be stabilized yet.

ABSTRACTOR'S ANALYSIS

The following comments and suggestions about the methodology and findings of this investigation are made. It is regrettable that the research design could not have included a control group and that the sample could not have been randomly selected. Because of this lack, any generalization of the findings of this research beyond the sample used in the study must be approached with caution. One wishes that there was more information in the report about the validity of the instrument used to measure attitudes. Data on the results of the administration of the instrument to the group of university professors would be helpful. Also, the low reliability of the instrument (.62) for the sample raises further doubts about generalizing the findings.

One of the variables is variously referred to as "number of years of college education," "level of education," and "university graduate" versus "nonuniversity graduate." It is difficult for the reader to determine what was actually measured for this variable because of this inconsistency. In Table V on page 252 there is an error in the recording of the difference between the pretest score (16.94) and the posttest score (16.59) for the university graduates. The difference is reported as .1 whereas, if the pretest and posttest scores are correct, the difference should be .35. This increases the value of t from .4 to 1.4. This new value of t is still not significant and the findings and conclusions remain the same.

An interesting question arises when one compares the finding that the four-week training course was ineffective in changing attitudes to the finding that prior professional inservice training was positively related to a change in attitudes. Why is prior inservice training related to a change in attitudes when this most recent inservice training is not? Perhaps the changes are cumulative or related to another variable not measured in this research, such as age. These questions might be an area of further research. The finding that teaching experience was not related to a change in attitudes is also interesting when compared to the positive relatationship found for inservice training: One might expect a relationship between these two variables since those with the greatest number of years of teaching experience would seem most likely to also have the larger amounts of inservice training. The question that arises here is how many in the sample had prior inservice training.

Perhaps the most interesting finding is the significant negative relationship between the level of education and the teachers' changes in attitudes. The investigators point out that a similar result was obtained by Hughes in a study involving elementary teacher trainees. They suggest that this result can be interpreted by assuming that university graduates have stabilized their attitudes toward science in comparison to nonuniversity graduates who have not yet stabilized their attitudes. Those with the least amount of education may have the most to gain. This finding can have important implications for the selection of participants for similar inservice training.

One can, of course, suggest alternate explanations of this finding. Again, age of the subjects seems a possibility. Are the university graduates older or younger than the nonuniversity graduates?

In summary, some of the findings of this research should be investigated further. Caution should be used in generalizing the results of this study because of some weaknesses in methodology.

Willson, Victor L. and Antoine M. Garibaldi. "The Association Between Teacher Participation in NSF Institutes and Student Achievement." Journal of Research in Science Teaching, 13(5):431-439, 1976.
Descriptors--*Achievement; Educational Research; Inservice Education; *Inservice Teacher Education; Science Education; *Science Institutes; *Secondary School Science; *Teacher Education; Teacher Improvement

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Edward J. Davis, University of Georgia.

Purpose

This study was directed at the question, "Is there any evidence that pre-college student cognitive achievement has been increased because of teacher participation in NSF-sponsored institutes?" It should be noted that this study was conducted as a NSF-sponsored project.

Rationale

The authors make the following argument for a post hoc analysis:

An experimental comparison between students whose teachers had attended institutes and students whose teachers had not would be optimal. The experiment would require random assignment of teachers to institutes (or not), and random assignments of students to teachers. Since NSF has not followed such a strategy, post hoc comparisons may be confounded by certain demographic and personological differences between teachers who have attended and those who have not attended NSF institutes. All potential factors can never be discounted in a post hoc analysis, but those theoretically most relevant should be dealt with (p. 431).

Research Design and Procedure

A post hoc analysis was performed. The authors identified science (or mathematics) achievement of teachers, and the level of classes to which a teacher is assigned to be the theoretical and relevant threats to examining the relationship between teacher institute attendance and students' academic improvement.

An urban-rural sample of junior and senior high schools was selected for science from Wyoming, South Dakota, and Mississippi and for mathematics from California and Indiana. Urban representation was small. Eighty-one percent of the science and 91 percent of the mathematics classes and teachers came from small towns and cities under 50,000 population. Within each school the principal was asked to select randomly one science (or mathematics) teacher and then select randomly one class from this teacher's load. This yielded a total of 346 science teachers and their classes and 211 mathematics teachers and their classes. Each teacher was given an achievement test in the subject area (NTE exams in either Physics-Chemistry-Science or Mathematics). Science students took a 40-item test taken from the NAEP science test and the mathematics students were given 40 items from the NLSMA item pool. Different 40-item forms were developed for junior high and senior high classes. Not all students took these achievement tests. Each teacher was given instructions to assign randomly attitude, process, and achievement instruments.

From a background questionnaire teachers were classified as having NO, LOW (1 or 2 institutes attended), or HIGH levels of participation in NSF institutes. This placed 36, 36, and 28 percent of the science teachers and 43, 29, and 28 percent of the mathematics teachers in the respective groups.

The procedures above provided the investigators with a means to control teacher achievement and level of class assignments which were identified as obstacles to examining NSF institute participation and student achievement. Teacher achievement on the NTE exams was used as a covariate in analysis of student achievement. The random selection of teachers and classes was used to produce a situation wherein approximately equal high-, middle-, proportions of high-, middle-, and low-ability classes appeared in the NO-LOW-HIGH partition of the teachers. The authors state:

> The possible differential assignment of institute attenders to higher-ability classes was examined by testing the independence

> of NSF participation from the teachers' assessment of the ability group of the class from which the achievement data were drawn (high ability, average ability, low ability, and mixed ability groupings). Also tested within the senior high school science data was the independence of type of class (biology, chemistry, and physics) from NSF participation. The chi-square statistic was used for each test, . . . All chi-square statistics were nonsignificant at p = .05, indicating independence of the distribution of teacher assignments by ability grouping, or subject matter in science, from NSF institute participation (p. 433).

Findings

It was reported that:

> The marginal means of student achievement for NSF participation show a consistent trend in the direction of better student performance with increased teacher NSF participation for all four analyses . . These means are essentially unaffected by adjustment for the covariate, since none of the regressions are significant at p = .10.... The nonsignificance of the covariate implies that teachers' science ability is not related with their students' achievement (p. 435).

To follow up differences in mean scores, two planned orthogonal contrasts were performed on the senior high science scores and two more on senior high mathematics scores. These contrasts used an F statistic. The first considered the combined scores of students of LOW and HIGH vs. NO teacher-institute participation. The second contrast compared the scores of LOW vs. HIGH participation. Three of these four contrasts had significance at the .01 level. These are reported as suggesting that teacher attendance at institutes is associated with higher student performance than no attendance, and that students whose teachers attended the higher number of institutes (more than 2) did better than students of teachers attending only 1 or 2 institutes.

Conclusions

The authors conclude that a real institute effect is present. They prescribe that institute attendance be required of all secondary science and mathematics teachers.

ABSTRACTOR'S ANALYSIS

This study investigates an important area. In terms of time and money, a great deal is being and has been invested in in-service education. Student achievement is seldom used as a criterion to evaluate in-service programs. It is relevant to do so.

I am left with some questions, however. When principals are contacted is it likely that they will select a science or mathematics teacher (and one of their classes) at random? Or will a principal tend to choose a teacher and a class according to some preconceived criteria in spite of guarantees of non-identification of participants? What about the levels of difficulty of the achievement tests? Were they constructed to reflect the range of cognitive behaviors identified in the NSLMA study (Computation-Comprehension-Application-Analysis)? What about the attitude and process measures? How were they constructed? How did the students perform on them?

Were these my only concerns, I would feel good about this study. However, I must take exception to the authors' conclusions and recommendations. The trend is for students having teachers who participated in NSF-sponsored institutes to have a significantly higher mean score than students having teachers who did not attend institutes. But how much higher are these means? About 1 or 2 points (items) on one 40-item test. With a large sample it is possible for such a small mean difference to be significant. Statistical significance is present but it is questionable whether this difference is meaningful or possessing any practical significance. When one considers the cost of an institute, to both sponsors and participants, a recommendation that teachers be required to attend them, based primarily on gains of 1 or 2 points on one 40-item instrument seems at best premature and at its worst feathering one's nest.

RESPONSES TO ANALYSES

A RESPONSE TO THE ANALYSIS OF

Mayer, Victor J., John Disinger and Arthur L. White. "Evaluation of an Inservice Program for Earth Science Teachers," by H. A. Smith. Investigations in Science Education, 2(4):12-14, 1976.

by

Victor J. Mayer and Arthur L. White

H. A. Smith, in reviewing the article by Mayer, Disinger and White in Volume 2, No. 4 (1976), makes several points that we feel are open to challenge. He suggests that the study is weak conceptually and the results obtained could have been predicted a priori. The study was an evaluation of a combined summer and inservice program to update teachers' knowledge and classroom skills. It is therefore difficult to understand what Smith means by conceptually weak. Perhaps he fails to distinguish between a research study which should have some underpinning of theory and evaluation which seeks to demonstrate the relative effectiveness of a program. The program evaluated had a consistent thread or focus. The instruments used were selected or developed because they reflected the content and philosophy of the ESCP program and/or the prevailing philosophy underlying modern junior high science curricula. The instruments selected therefore were consistent with the objectives of the program and the design should reflect the underlying philosophy of the program.

In stating that the results could have been predicted a priori he places a great deal of confidence in the ability of the program developer or he misses the objective of an evaluation program. The example used in support of his statement is that one would logically expect teachers to gain geological knowledge from a four-week exposure to full-time instruction. If the program is a good one and the teachers are motivated properly, this should indeed by true. The purpose of an evaluation is to see if in fact it happened. In point of fact it did not happen with two teachers; one demonstrated a negative gain and one a very slight positive gain in knowledge related to the objectives of the ESCP.

His comments that the analyses do not separate curriculum and teacher effects (we assume he means teacher training effects) are valid. This could not be done because the population was too small to run the necessary analysis and the resources available (as is often the case in evaluation studies) were not sufficient to set up the necessary comparison groups. This from our point of view is the major deficiency in the study. Were this study to be done again, in the light of the increasing maturity of the field of educational evaluation (this study was started almost 10 years ago), a much greater effort would be made to identify a larger number of teachers so that a comparison group (or groups) would be available and to distinguish between those using ESCP and traditional approaches during the three-year term of the study. Concurrent with this, an attempt would have to be made to obtain the necessary funding.

We concur in the disappointment expressed by Smith in the results of the assessment of student achievement during the first year of the study. It reinforces the need for such evaluation studies and contradicts his statement that results could have been predicted a priori. If we were to attempt a priori predictions we would conclude that with such better-prepared teachers the students would show significant gains in science concepts (stressed by the ESCP) as well as in facts. This was clearly not the case. Following the second year, however, students did show significant gains in both understanding of concepts and of processes of science. It may be that prolonged contact between program staff and teachers is necessary before the program has its effect upon curriculum. It is disappointing that Smith apparently did not read far enough in the article to come across the report of these second year gains. It is these gains that the conclusions of the study are based upon. Not the first year gains, or lack thereof.

Smith contests the use of teachers as N rather than students. Since teachers, their training and consequent influence on the classroom are the focus of the study (teachers were enrolled in the inservice program, not students) it is indeed the teacher that is the appropriate N. If students had been used, as is common and inappropriately done in many studies in science education, all student gains (even the first year concept attainment) would have been insignificant, merely because of the much higher N. In this study, because of the nature of the program, the appropriateness of the teacher as N is obvious. We would submit, however, that because of the uniqueness of a class environment, all studies that involve learning in a class situation should use the class (and perhaps even the teacher in some cases) as N. Even students in different classes of the same teacher are exposed to totally different learning environments. Research and evaluation designs must begin to recognize these differences and account for them.

In summary, we are disappointed that the analysis of the study seems so superficial. He seems to view it from the arm chair researcher's point of view rather than that of the practicing evaluator. This study is only one of two published studies that we know of that has attempted a longitudinal evaluation of a science inservice program and the only one extending beyond the termination of the program. The review seems to have entirely missed this point. It is hoped that published studies, even if imperfect like the one discussed here, will spur others to conduct similar but improved studies and report on their results. Only when the results of many such studies are available will we be able to document the efficacy of different approaches to inservice education. To us that is the importance of this study and we are distressed that it was not identified as such by the reviewer.