DOCUMENT RESUME

ED 113 187 SE 019 732

AUTHOR	Helgeson, Stanley L., Ed.; Blosser, Patricia E., Ed.
TITLE	Investigations in Science Education, Vol. 1, No. 3. Expanded Abstracts and Critical Analyses of Recent Research.
INSTITUTION	Ohio State Univ., Columbus. Center for Science and Mathematics Education.
PUB DATE	75
NOTE	71p.
AVAILABLE FROM	Ohio State University, Center for Science and Mathematics Education, 244 Arps Hall, Columbus, Ohio 43210 (Subscription, $6.00, $1.75 single copy)
EDRS PRICE	MF-$0.76 HC-$3.32 Plus Postage
DESCRIPTORS	*Abstracts; *Educational Research; Research; Researchers; *Research Methodology; *Research Skills; *Science Education

ABSTRACT

This third issue of INVESTIGATIONS IN SCIENCE EDUCATION (ISE), designed to provide a new perspective for viewing research articles and to aid in the improvement of writing research reports, includes abstracts prepared by science educators, bibliographical data, purpose, rationale, research design and procedure, findings and interpretations as well as detailed notes offered by the abstractor. Articles included in ISE are selected primarily from such sources as professional journals and reports of government-funded projects. Abstracts in this issue relate to such topic areas as evaluation of elementary science curricula utilizing behavioral tasks, effectiveness of audio-tutorial instruction, effectiveness of tutorial computer-assisted instruction, effect of verbalizers on achievement of non-verbalizers in an enquiry classroom, and a study of verbal behavior patterns in primary grade classrooms during science activities. (EB)

Vol. 1, No. 3 1974-75

INVES
IN
SCIEN
EDUC

INVESTIGATIONS IN SCIENCE EDUCATION

Expanded Abstracts and Critical Analyses of Recent Research

National Association for Research in Science Teaching

ERIC Science, Mathematics, and Environmental Education Information Analysis Center

The Center for Science and Mathematics Education
The Ohio State University

INVESTIGATIONS IN SCIENCE EDUCATION

National Association for Research in Science Teaching

ERIC Science, Mathematics, and Environmental
Education Information Analysis Center

Published Quarterly by

The Center for Science and Mathematics 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.

Vol. 1, No. 3, 1974-75

NOTES . . .

from the Editor

Beginning with Vol. 1, No. 4, Investigations in Science Education will reflect some changes suggested by the ISE Advisory Board. One of these involves an attempt to "cluster" analyses of research reports around a central theme, topic, or investigative approach. This will permit a more in-depth examination of current research in science education focusing on the underlying theoretical framework. Another proposal includes the analyses of some research reports from differing points of view (in-paradigm/out-of-paradigm). Thus, a Piagetian study might be reviewed by an advocate of Piaget's views of cognitive development and by an individual holding contrasting views. These changes are not intended to supplant the review of individual reports of research, but rather to supplement them.

Again, we encourage your comments and suggestions for improving Investigations in Science Education.

Stanley L. Helgeson
Editor

Patricia E. Blosser
Associate Editor

EJ 068 807

Alexander, Donald L. and Alan C. Donaldson, "Earth Science in West Virginia Schools." Journal of Geological Education, Vol. 20, No. 4:193-195, September, 1972.
Descriptors--Curriculum, *Earth Science, *Educational Programs, Geology, Research, Science Education, *Secondary School Science, Surveys

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Victor J. Mayer, The Ohio State University.

Purpose

The stated purpose of the study was to determine the current status of earth science as a curriculum offering in the secondary schools of the state of West Virginia. It was implied, but not stated explicitly, that the study was carried out in 1971.

Rationale

The introduction to the article implies that the junior author conducted a similar study in 1957, reported in 1963, in which he found that only one West Virginia school offered earth science. The present study was initiated to determine the progress in school adoptions of earth science curricula in the 14 years following that first study.

Research Design and Procedure

The study was conducted in two stages. The authors first obtained a status report on earth science offerings from the Department of Education of the state of West Virginia. They provide no indication as to the nature and source of data for this report nor how the information was compiled. One must assume that it did provide the names of those teachers in the state who were teaching earth science.

In the second stage more detailed information was obtained through the use of a questionnaire sent to every teacher of earth science in the state. The questionnaire was designed to assess:

a) pupil attitude toward earth science;
b) teacher attitude toward earth science;
c) teacher preparation in earth science;
d) materials being used by the teacher in the classroom;
e) methods being used by the teacher in his earth science classes; and
f) a rationale for the continued growth of earth science in West Virginia as expressed by the earth science teachers.

The questionnaire was sent to 107 teachers apparently identified by the State Department of Education as teachers of earth science. Responses were received from either 40 or 42 teachers (both figures are cited at different places in the article), representing 39.6

percent of the population. Apparently no follow-up was attempted to gain greater response, nor was there any attempt to determine whether the non-respondents differed in any way from those teachers who did respond.

The analysis consisted of tabulation of the data and conversion to percentages.

Findings

The major findings as cited by the authors are as follows:

1. The number of schools offering earth science in the state increased from one school in 1957 to 73 schools when the most recent survey was conducted (assumed to be 1970-71).

2. Earth science is most commonly taught in grades 7-9, with the highest number of students enrolled in the course in grade 9.

3. Only four of the 40 respondents were teaching earth science full-time. The remaining teachers taught other sciences, English, music or driver education in addition to earth science.

4. The response to the items on teacher background revealed that 65 percent were certified to teach earth science. Twenty percent were not sure whether they were certified to teach earth science. The authors did not provide information as to the nature of the state certification requirements. The number of years teaching earth science averaged 3.8. The most frequently indicated science area taken by the teachers was physical science, averaging 7 credits per teacher. Eleven teachers indicated no background in it. Geology was also taken by all but 11 of the teachers, with the average number of hours being 5. The hours of an area taken were averaged over the entire number of teachers responding, rather than on the basis of the number of teachers indicating having background in that area. This gives a rather misleading impression of the depth of background in each area. Also, it is not indicated whether the hours are expressed in semester or quarter hours of credit.

   Of those teachers returning the questionnaire, 65 percent indicated that they would like to take additional background in earth science; 60 percent wanted more geological field training, an area which the authors state is neglected in West Virginia. The teachers would prefer summer institutes to other forms of additional training (50 percent responding favorably to that item on the questionnaire).

5. Teachers were also asked to list the materials that they used in teaching earth science. The most frequently used textbook was Modern Earth Science by Ramsey and Burkley (46.3 percent).

The Earth Science Curriculum Project (ESCP) textbook, Investigating the Earth was used by only three teachers (7.3 percent). The most frequently used equipment was film and filmstrips (60 percent); topographic maps were used by 35 percent of the teachers; only 2 (5 percent) used the ESCP laboratory equipment. Fifty-five percent of the teachers did not use laboratory sessions as an integral part of their earth science course. An explanation is suggested, according to the authors, by the fact that 35 percent of the teachers felt that their laboratory facilities were "below average" while 62.5 percent responded that their laboratory facilities were average or above average. Only 37.5 percent of the teachers indicated that field trips were integral parts of their courses. They conducted an average of two per year.

6. Of the teachers responding to the questionnaire, 97.5 percent felt that earth science offerings in the state should be expanded. They felt that a knowledge of earth science was important in: "(1) environmental understanding; (2) satisfaction of children's curiosity of local field area understanding; (3) intellectual value; (4) training in using natural resources wisely; (5) understanding the effects of pollution; and (6) acquainting children with our home, Earth." The authors did not indicate if these reasons were the result of a free response option on the questionnaire or of rankings provided from forced choice items. There is also no indication how frequently each of these reasons was offered.

Interpretations

The authors conclude that there need to be some drastic changes in the way earth science is being taught in West Virginia, in light of the limited degree to which laboratory and field experiences are being offered by the teachers responding to the survey. They suggest that a possible answer to this problem would be to offer more field training for teachers. They also conclude that earth science must be made available to more students in the state, since "only 4.9 percent of the public school students are currently exposed to an earth science course." In addition, improvement is needed in the physical facilities for earth science laboratories and in teacher preparation.

Abstractor's Notes

The study reported in this article should be a valuable source of information for educational planners in the state of West Virginia. It is also of interest to those in earth science education around the country in that it documents the growth of earth science curriculum offerings in a state where problems associated with the utilization of earth resources are in the forefront of national attention. The information would appear to be valid and fairly well documented. The report of the study, however, is replete with problems in interpretation, manner of statement, citation of data and vagueness of relationships between conclusions and data. In general there is a failure to supply enough information to allow the reader to evaluate the conclusions which are drawn by the authors.

The authors, in the introduction to the article and in a table on the same page, cite statistics on earth science enrollments in 1957 and again in 1962-63. They do not supply any information on how these statistics were gathered in the study being discussed. In the next section (entitled "Current Status"), it is stated that information on current status was obtained from the West Virginia State Department of Education, and the reader is left to assume that the information in the previous two studies came from the same source. If this is true, then, hopefully, comparisons can be made, provided the state department has not changed its methods of acquiring and reporting school information in the interim. The authors also fail to state the procedure used by the State Department of Education in acquiring its information. If it is based on annual reports from principals, as it is in Ohio, it could be highly reliable, but the reader does not know. In the same section a statement is made that of 55 counties in West Virginia, 30 offer earth science. This would imply that West Virginia, unlike the rest of the country, operates on a county-wide school system. What the authors probably meant to say is that earth science is offered by schools locat d in 30 different counties.

The only major fault with the design is an apparent failure to pilot the questionnaire and to determine its reliability or validity. This is an unfortunate characteristic of many questionnaire studies. Further compounding the problem in this particular article is the failure to state the questions used in the questionnaire. This could very readily have been done early in each section of the article relating to a particular question. This would give the reader the opportunity of judging the validity of each question and therefore the quality of the information the question generated. Apparently the authors failed to send out a follow-up questionnaire. Doing so could have increased their return substantially. In any case, they failed to determine if there were any substantial differences between the respondents and the non-respondents. This could readily have been done by randomly selecting ten or so non-respondents, obtaining their response to the questions over the telephone, and then comparing their responses with the others. If there were no great differences, than the conclusions of the study could be generalized for the entire population of West Virginia earth science teachers. As it is, the authors consistently conclude that the results of this survey represent the status of earth science in the state as a whole. While this may be true, the design does not permit such generalizations.

There are numerous statements that bother the careful reader of the article. At the end of the section entitled "Survey Questionnaire," the statement is made that "The total population of 107 earth science teachers in West Virginia was sampled." What is meant is that the questionnaire was sent to the entire population of earth science teachers in the state. In discussing the preparation of earth science teachers as indicated by the survey, the statement is made that "The preparation of teachers in earth science is still far behind the preparation of teachers in other areas of science." One must assume that what is meant is that earth science teachers have less background in earth science courses than other science teachers have in their particular disciplines. Even if one makes this assumption, however, it is still not clear if the statement means other science teachers in West Virginia as the referent group, or other science teachers nationwide. What is even more serious, the authors fail to cite the source of their knowledge on the preparation of other science teachers. A similar problem occurs in the section entitled "Methods and Materials used in Teaching Earth Science." In discussing the failure of earth science teachers to include laboratory work in their courses, they state that "Marked decreases in learning and in the desire to learn take place when active student involvement is not utilized in science courses." No studies are cited which might lend support to such a statement. In fact, there appears to be little empirical support for this statement. Subsequently the authors attempt to explain the failure to use laboratory activities partly upon the teachers' perceptions of their laboratory facilities. This is based upon a question in which 35 percent of the respondents rated their facilities below average, and the remainder rated them average or above. One of the "advantages" of earth science as a laboratory course is that it requires very minimal laboratory facilities. It is here, perhaps, where the bias of the authors begins to show through, especially when concluding that only 37.5 percent of the teachers were conducting an average of two field trips per year as an integral part of their course.

One final error that is quite disturbing is found in the authors' conclusions where they state that only 4.9 percent of the public school students are currently exposed to an earth science course. This figure is apparently based upon the statement made at the beginning of the article that 4.6 percent of the students in grades 7-12 were taking earth science in the year that the report of the State Department of Education was compiled. The authors apparently made an error in quoting the percentage. If we assume that a student would take earth science only once in secondary school, ignoring repetitions due to failure and the availability of second level earth science courses, and that the relative proportions of students in the various grade levels remain constant between school districts throughout the state, then the maximum probability that a student would take earth science is: 6 (the six years that a student would be enrolled in secondary school grade levels) times 0.046 (the probability that a student takes earth science in a given year) or 0.276. In other words, one would expect a maximum of 27.6 percent of the graduates of high schools in the state of West Virginia to have taken an earth science course while in the secondary schools.

EJ 066 421

Atwood, Ronald K., Mark R. Brown and Anna A. Neal, "Evaluation of a Hybrid Elementary Science Curriculum Utilizing Behavioral Tasks." School Science and Mathematics, Vol. 72, No. 7:641-646, October, 1972.

Descriptors--*Behavioral Objectives, *Curriculum Evaluation, *Discovery Processes, Educational Research, Elementary Grades, *Elementary School Science, Evaluation, Instruction, Science Education

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Glen O. Blough, University of Maryland.

Purpose

To describe an initial effort to develop a scheme that utilizes behavioral tasks for the evaluation of an experimental elementary science curriculum in a school with modest resources. Hypotheses tested for group post test means were:

1. The experimental primary group will score significantly higher (.05 level) than the control primary group

2. The experimental intermediate group will score significantly higher (.05 level) than the control intermediate group.

3. The total experimental group will score significantly higher (.05 level) than the total control group.

Rationale

The twenty behavioral tasks for the primary level (readiness through grade 3) were selected from the SAPA Competency Measures with few and minor modifications. The processes included and the number of tasks used to evaluate the processes were: observing (3), classifying (3), using numbers (2), space-time relations (3), communicating (4), measuring (4), and predicting (1).

For the intermediate level (grades 4 through 6) fifteen tasks were selected from the SAPA Competency Measures with few and minor modifications, and five similar tasks were prepared by the investigators. The processes included and the number of tasks used to evaluate the processes were: observing (2), classifying (2), using numbers (2), space-time relations (1), communicating (3), measuring (5), predicting (1), inferring (1), control of variables (1), and design of experiments (2). The last ten tasks used with the primary group were identical to the first ten tasks used with the intermediate group. The philosophy and general objectives of the school system provided the basis for task selection.

Research Design and Procedure

The Fayette County Schools, Lexington, Kentucky, decided in the Fall of 1969 to pilot Science, A Process Approach (SAPA) and Elementary Science Study (ESS). Because a "whole-school" approach was preferred and funds were rather limited, SAPA was implemented at the primary level and ESS at the intermediate level in a single small urban elementary school. A school comparable in terms of student socio-economic levels, enrollment, racial balance, curricula (other than science), and teacher preparation was available to serve as a control. The science curriculum in the control school relied heavily on a content-centered text.

Since the philosophy for science instruction officially adopted by the Board of Education clearly favored process-oriented education, as opposed to information storage process-oriented tasks were chosen as the most meaningful mode of measuring student performance.

Recognizing task administration on an individual basis to be very time consuming, a random sample of approximately 20 percent of the school enrollment was chosen by grade level in both experimental and control schools. One investigator administered the tasks, both pre and post, to all primary level subjects. Another investigator administered the tasks, both pre and post, to all intermediate level subjects. Performance on each task was scored as either acceptable or non-acceptable, based on specific criteria, and each student's score represented his number of acceptable responses. Pretesting was done during early October, 1969, and the post administration was completed in early May of 1970.

The teaching staff of the experimental school received six days of special preparation prior to the opening of school, and three days during the school year. Teacher Corps interns and special education faculty participated. The control school faculty, which also included Teacher Corps interns and special education faculty, received no preparatory or in-service experiences in elementary school science.

Findings

Two tables supply the data results. Table I is a data summary. The treatment means indicate the arithmetic averages of the raw scores, the adjusted means were the result of post test scores with the pre test scores used as a covariate. In Table II the adjusted means were used in analysis of covariance to test the hypotheses.

TABLE I. DATA SUMMARY

Group	Level		Treatment Mean	Adjusted Mean	n
Experimental	Primary	Pre	5.48		21
		Post	9.14	8.97	21
	Intermediate	Pre	3.61		23
		Post	6.39	6.85	23
	Primary & Intermediate	Pre	4.50		44
		Post	7.70	7.85	44
Control	Primary	Pre	5.00		20
		Post	5.20	5.38	20
	Intermediate	Pre	4.72		25
		Post	5.72	5.30	25
	Primary & Intermediate	Pre	5.84		45
		Post	5.49	5.35	45

TABLE II. ANALYSIS OF COVARIANCE SUMMARY

Groups	Source	df	Mean-Square	F	P
Primary	Treatment	1	130.68	44.50	.01
	Error	38	2.94		
Intermediate	Treatment	1	27.00	5.99	.05
	Error	45	4.51		
Primary & Intermediate	Treatment	1	137.97	35.04	.01
	Error	86	3.94		

On the basis of these data the experimenters concluded that the experimental groups out-performed the control groups on the behavioral tasks and all the hypotheses were accepted. They caution about interpreting the difference of the levels of significance for the primary and intermediate groups since only ten of the twenty tasks were common to both levels and comparisons made involved only experimental vs. control groups.

The experimenters point out that the difficulty level of the tasks, of initial concern to the investigators, proved to be acceptable. All students scored above zero on pre and post administrations, with a total of five students scoring one. A total of five scores above 12 were recorded with a single high score of 16 most closely approaching the maximum score of 20. Thus, "topping out" was not a problem in this study and would likely not become a problem were the same tasks utilized in a longitudinal study extending another year.

Examination of results for each individual task provided useful information. For example, on one task primary level subjects were asked to make a group or set of all the large blue triangles after they had been presented with a set of the ESS "A" Blocks. On the pre-test eight subjects in each group demonstrated competency while on the post-test 15 subjects in the experimental group and only seven subjects in the control group demonstrated competency.

They further point out that from some tasks plus observations made while the tasks were being administered, it was inferred that the language used to present the tasks masked the extent to which some process skills were developed and, further, that this masking effect was greater for the control group.

Findings and Interpretations

The experimenters indicated that the assumption that behavioral tasks provide a sound mechanism for assessing curricular outcomes judged to be especially important at the local level is supported by this study. Comparisons of experimental and control groups produced statistically significant results in favor of the experimental groups. Subjective examination of results by task and clusters of tasks were considered fruitful. To strengthen these subjective analyses one could use chi square where cell frequencies are adequate. One of the real bonuses of utilizing behavioral tasks lies in the insights individuals administering the tasks gain in how children perceive and attack task solutions.

One of the greatest sources of difficulty is the language of the task, which can mask skill development. A more obvious difficulty rests with attempts to consistently administer the tasks.

It should also be remembered that failure to demonstrate competency on a particular task provides limited information, since subordinate (enroute) behaviors may have been developed. Not knowing the extent to which this may have occurred is an unavoidable problem when any sample of tasks is utilized, but the problem can be expected to become larger as the sample of tasks becomes smaller.

Abstractor's Notes

While such research as this adds to the significant knowledge of methods and procedures in science teaching and learning, a compiling and comparison of the results of similar studies seems highly desirable before truly dependable results can accrue.

In learning studies, as we all know, there are many variables that are difficult to control effectively. For example, children in the experimental group may indeed respond positively to the situation in which special teachers are involved and because of this achieve to their greatest capacity while the routine in the control group does not elicit such effort. It is also common knowledge that without extra training the teachers in the control group cannot possibly measure up to the standard of the investigators. In many cases the teachers of the control groups are not only untrained in the methods of instruction, they may indeed be unaware of the objectives for teaching the material while the experimental group instructors are teaching with specific objectives in mind. No amount of statistical manipulation can really compensate for these differences.

What may be thought of as a byproduct of such investigation is the growth and development made by the investigators in understanding the many facets of how children learn, their difficulties, potential, the importance of motivation; indeed, an understanding of the problems involved in teaching at the elementary school level. The experience may also result in a more thorough understanding of how science fits into the total elementary school program. It may further result in examining some of the behavioral tasks and processes in light of their appropriateness and significance of young children. As is the case in many of our investigations into the learning of children, careful study of the teaching process as it goes on and the data that result cause us to proceed with caution in forming generalizations from our experience and results. Many of the so-called studies in accountability presently in process will do well to apply a more scientific attitude toward interpretation.

EJ 042 718

Hoffman, Frederic and Marvin Druger. "Relative Effectiveness of Two Methods of Audio-Tutorial Instruction in Biology." Journal of Research in Science Teaching, Vol. 8, No. 2:149-156, 1971. Descriptors--*Audiovisual Instruction, *Autoinstructional Aids, College Science, Educational Strategies, *Genetics, *Instruction, Research

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Martin Hetherington, Michigan State University.

Purpose

The purpose of this research was to analyze and compare the relative effectiveness of two different audiotape strategies with respect to their effect on student achievement, student retention, attitudes, problem solving abilities and critical thinking abilities. There were seven hypotheses tested.

Rationale

Since the inception of audio-tutorial instruction, no systematic attempt has been made to analyze and compare the relative effectiveness of different audiotape strategies. In the past many of the audio-tutorial programs have been structured intuitively or through trial and error by the instructor.

Research Design and Procedure

A sample of 90 students, 43 males and 47 females, was randomly selected from more than 800 students enrolling for the first semester of a general biology course at Syracuse University. Only five of the students had prior experience with audiotaped instruction. The sample was randomly divided into two equal sized groups, each being taught by a different audiotape strategy in separate audiotape laboratories. Each laboratory was equipped in the same way. Students were provided with guidesheets for the current lesson when they entered the laboratory. These guidesheets contained charts, diagrams, problems and instructions for laboratory work.

At the beginning of each week the students signed up for as many two hour blocks of laboratory time as they wished for that week's work. There were no discussion periods or lectures. General information and quizzes were administered once a week.

The experimental period lasted six weeks. The subject matter was the same for both groups: mitosis, meiosis, and genetics. The major difference between the two experimental groups was the audiotape teaching strategy.

The direct group received lessons of a descriptive or expository nature that usually involved passive notetaking by the students. The emphasis was on listening, with little or no question and answer activity. Students were directed to examine diagrams, film loops, slides and guidesheets. There were no self-quizzes and very little problem solving or investigations done by the students in this group. The laboratory work was guided by audiotapes and written instructions, with all observations, conclusions and answers to questions stated on guidesheets. The amount of student-teacher interaction was minimal in this method.

The indirect group was taught by a lecture-question-answer method in which the students proceeded toward the objectives in an investigatory manner. Lecturing was on the essential information and no more. The film loops, slides, and demonstration materials were primarily the same as those used for the direct group. The guidesheets were different in that they contained only a skeletal framework as a guide to learning. The students used the guidesheets to develop answers to questions, form their own conclusions and solve problems. Laboratory assignments were outlined on the guidesheets and the students answered questions and made observations to accomplish the laboratory objectives. The amount of student-teacher interaction within the taped lessons was much greater in this group than in the direct group.

An analysis instrument had to be developed in order to determine whether the two strategies used were different. This instrument was similar to that used by Flanders for classroom interaction. Nine categories of classification were used; four categories were directed at the Indirect group and five categories, at the Direct group. The first four categories of classification, 1 through 4, were designed for the Indirect group and came under the heading of "Motivating Statements." The last five categories were designed for the Direct group. Categories 5 and 6 were under the heading of "Information Giving," and 7, 8, and 9 under the heading of "Control Statements." The taped lessons and the laboratory exercises were analyzed using this instrument.

The analysis instrument was designed to determine the directness or indirectness of a lesson based on teacher input as obtained from tape manuscripts and guidesheets for each group. The categories were grouped into direct and indirect activities, according to the freedom of action of the student during the taped sequence. Direct activities were either of an information giving nature or a form of control. Indirect activities were those in which the student's freedom of action was considerably increased and in which he may have been mentally motivated or encouraged. Each taped script and laboratory guidesheet was evaluated according to the nine categories. The analysis was reached in sequence as the categories occurred in the scripts. Category numbers were then compiled in a matrix, one pair at a time.

To determine the reliability of the analysis tool, a Scott's correlation was computed of each of 10 taped lessons using three self-trained observers.

The Watson-Glaser Critical Thinking Exam, form Ym, was used to measure critical thinking changes and transfer of learning abilities.

A subject matter exam, prepared by the investigator, was used to test differences between the two experimental groups in achievement, retention, and problem solving abilities. The reliability of the 60-item exam was determined to be 0.86 using the Kuder Richardson formula 20. The content validity was maintained by including a cross-sectional sample of a large number of items representing the areas in which performance was to be evaluated. The test items were developed by instructors in the field of genetics. An item analysis based on 175 students taking each of the 60 multiple choice items was made.

A further breakdown of the 60 items was made to derive the two types of questions used in this investigation. The knowledge type questions were intended to measure the student's ability to recall facts and make simple observations. The use of knowledge type questions was designed to test the student's ability to apply his knowledge and solve problems, state and make judgements. The 60 questions were reviewed by a panel of five educators to determine the appropriate type.

An attitude questionnaire was prepared by the investigator to evaluate attitude changes in the two groups.

Each test instrument was given to both groups. Both groups were pretested and posttested.

Findings

The investigators found that they could train observers to use their analysis instrument with a high level of agreement in their script and laboratory analyses.

They were successful in constructing a 60-item multiple choice test which had the following discriminating powers: 55 percent were 0.4 and over, 37 percent were 0.2 to 0.4 and 8 percent were 0.2 and under.

The attitude questionnaire, which was prepared by the investigation, indicated that the pretest scores would reflect the two groups to be homogeneous at the start of the course. All the "t" values which were computed to determine differences between the two groups fell well below the required t (.05, 85) of 1.99.

The seven null hypotheses tested were:

Null Hypothesis 1: At the conclusion of the experimental period, the two groups will not differe with respect to their mastery of facts, concepts, and principles concerning heredity as measured by a subject matter exam prepared by the investigator.

A t-test was performed to test differences between the means of the pretest scores of the two groups. That was determined to be 0.949 and the required t (.05, 85) was 1.99. The results of the t-test indicated that there was no significant difference between the direct and indirect groups with respect to their mastery of facts, concepts and principles concerning heredity. Therefore, the hypothesis was not rejected.

Null Hypothesis 2: At the conclusion of the experimental period, the two groups will not differ with respect to their problem solving ability as measured by the use of knowledge questions from the investigator's subject matter exam.

A t-test was performed to test differences between the means of the two groups with respect to the use of knowledge questions.

The t obtained was 3.03 and the required t (.05, 85) was 1.99. The results of the t-test indicate that there was a significant difference between the two groups. The indirect group performed significantly better, with respect to problem solving abilities. Therefore, the null hypothe is was rejected.

Null Hypothesis 3: The two experimental groups will not differ with respect to their ability to retain factual and conceptual information was measured by a subject matter test prepared by the investigator.

The observed t value for the t-test for difference between the means of the two groups was 1.07 and the required t (.05, 85) was 1.99. Based on the results of the t-test, the hypothesis was not rejected since there appeared to be no significant difference between the two groups with respect to their abilities to retain factual and conceptual information.

Null Hypothesis 4: At the conclusion of the experimental period, the two groups will not differ with respect to their ability to think critically and to transfer learning to other situations as measured by the Watson-Glaser Critical Thinking Exam, form Ym.

The t-test results for differences between the means of the two groups produced a t of 0.18. The required t (0.5, 85) for a significant difference was 1.99. Based on the test, there appeared to be no significant difference between the critical thinking abilities of the two groups, and therefore, the hypothesis was not rejected.

Null Hypotheses 5, 6, and 7: Three null hypotheses were tested with regard to the attitude questionnaire.

At the conclusion of the experimental period neither group will show a difference from their original attitude toward:

(a) biology and genetics as course subjects
(b) the indirect teaching strategy
(c) the audio-tutorial method of instruction

Each item on the questionnaire was assigned to one of the three categories and each was examined to determine the direction of the shift in the response patterns of the two groups from the pretest to the posttest.

Each of the hypotheses was examined in light of the following questions:

(1) Were there any individual items which showed significant changes between the pretest and the posttest for either group?

(2) If there were any significant changes, what were the directions of change; increase or decrease (positive or negative)?

(3) Was there a significant difference between the number of items showing a positive shift and the number of items showing a negative shift in attitude for either group?

Within the category of genetics and biology as course subjects, the two significant responses indicate that students of both groups preferred the genetics units at the beginning of the course. The one significant negative response indicated that the direct group felt that biology was a difficult course.

The indirect group felt that there was sufficient information on the guidesheets while the direct group felt that there should have been more. Neither group believed that the students should be made to search for knowledge, but they liked the idea of weekly quizzes.

In general, the students felt that the merits of audio-tutorial instruction far outweighed any shortcomings and that it was an effective and inspiring method of teaching. They felt that they learned more in a shorter time, were encouraged to think critically, were not frustrated by the method and were helped by the guidesheet and the clearly defined objectives.

Not all of the items on the questionnaire reflected a significant shift from the students' original attitudes.

Out of the 10 items related to biology and genetics as course subjects, both groups reflected 5 positive and 5 negative shifts. From the six items pertaining to the indirect strategy, both groups showed three positive and three negative shifts. Neither group reflected a significant change from their original attitudes. Therefore, the hypotheses concerning these two aspects were not rejected.

Out of the 28 items related to audio-tutorial instruction, the direct group showed 28 positive and 0 negative shifts and the indirect group showed 26 positive and 2 negative shifts. Therefore, this hypothesis was rejected since there was a difference from the students' original attitudes toward the audio-tutorial method.

Interpretations

The results of the study indicated that both the direct and the indirect audio-tutorial strategies were equally as effective for teaching facts, concepts, and principles concerning heredity. Both strategies had the same effect concerning the retention of facts, concepts, and principles concerning heredity. Both strategies resulted in a significant change in students' attitudes in favor of audio-tutorial instruction.

Neither the direct nor the indirect strategy improved students' critical thinking abilities. Neither caused a significant change in the students' attitude toward biology and genetics as course subjects or toward the indirect teaching strategy.

There was a significant difference between the two strategies with regard to problem solving abilities. The results of the study suggested that the indirect teaching strategy might be a more effective teaching strategy than the direct for developing problem solving abilities.

The tape analysis instrument developed during this study provided a means to objectively analyze the directness or indirectness of taped instruction. The results do not indicate highly significant differences between the direct and indirect teaching strategies.

Abstractor's Notes

Methods to quantify the value of audio-tutorial teaching have been needed for a long time. This paper tries to get at one technique to do this. The fact that the main finding of this research was the interest of students in the audio-tutorial teaching is not new. The survey points out that of the 90 students used in the research, only 5 had prior experience with audiotaped instructions and that this instruction was in language laboratories. This means that for most of the students the Hawthorne effect was a possibility. Similar research should be done with a student population who had been exposed to several audio-tutorial courses before the direct-indirect teaching strategies were applied.

Why wasn't the attitude questionnaire also given to include attitudes about the direct teaching strategy?

The results might have been expected to come out the way they did; this does not necessarily mean that this is the best way to teach problem solving to students.

EJ 042 717

Hughes, Earl F., "Role Playing as a Technique for Developing a Scientific Attitude in Elementary Teacher Trainees." Journal of Research in Science Teaching, Vol. 8, No. 2:113-122, 1971. Descriptors--Elementary School Teachers, *Instruction, Methods Courses, Research, *Role Playing, Science Education, *Scientific Attitudes, *Teacher Education

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

Purpose

The problem of devising an effective teaching technique for developing a more positive attitude toward science, scientific endeavors, and scientists as applied to preservice elementary teachers is the general focus of this study. The null hypotheses tested were:

(1) There is no significant difference on the science attitude measure between the control group and the combined treatment groups of reading a persuasive communication, role playing, and observing the role playing.

(2) There is no significant difference between the group that only reads the communication and the combined observer and role playing groups on the science attitude measure.

(3) There is no significant difference between the observer group and the role playing group on the science attitude measure.

Rationale

The author contends that elementary teachers should have a positive scientific attitude in order to more effectively teach elementary children an active, hands-on version of science. Since many preservice teachers have experienced science only as an organized body of knowledge, they have not been given the opportunity to formally experience the process and concept development components of science. As a result of this limited experience, preservice teachers in general do not exhibit a positive attitude toward science, scientists, and scientific endeavors.

A review of research studies related to existing characteristics of science teachers and science students provides a broad series of conclusions that directly or indirectly relate to the stated problem and hypotheses. Generally, elementary teachers are reluctant to teach science, the amount of science training for teachers is positively correlated to opinions held by scientists, and college science classes do not contribute to an understanding of science.

A fundamental assumption of this study is that college science courses appear to increase the preservice teacher's confidence and competence in science, but they do not cause the prospective teacher to teach for current curricular objectives at the elementary school level. The investigator assumed that an attitude change technique using persuasive communication and role playing would effect attitude change in preservice teachers as part of a science methods course. Six studies are cited that deal with attitude change techniques.

Research Design and Procedure

Four groups were formed. Group C, the control, read an innocuous communication and responded to ten science attitude items (post-test). Group R read the persuasive communication and responded to the science items. Group RO read the persuasive communication, listened to and observed the role playing, and responded to the science items. Group RRP read the persuasive communication, role played a proponent of the position of the communication, and responded to the science items. The persuasive treatment was administered four weeks after the pretest was given. The pretest scores were used to assign subjects to experimental groups with no significant difference between the mean science attitude of the four groups. Nine days after the treatment a post-post-test, identical to the pretest, was administered.

The sample of 184 was drawn from science methods or physical science for elementary teachers classes from four colleges. The students of an instructor from each institution were used. There were 28 males and 176 females in the 204 post-test subjects. Of these, 97 were juniors and 96 were seniors. Nine subjects had incomplete data on year in school.

The persuasive communication took a position on each of the ten questions included in the attitude measure. The main points emphasized concerned benefits from a discovery approach and science as a dynamic, analytical, flexible way of approaching a problem. The innocuous communication was an article on the characteristics of a good test. Both communications were approximately 2000 words long and required reading rates of 200 words per minute to complete in the allotted time.

The science attitude measure, post-test, used a seven choice Likert-type scale ranging from strongly agree to no opinion to strongly disagree.

The pretest and post-post-test were identical. The ten science related items used in the post-test were embedded at random within a 34 item opinionnaire including science, mathematics, social studies, and language arts questions.

A computer program using a multiple linear-regression analogy to a single-classification analysis of variance was used to test the three hypotheses.

Findings

Hypothesis 1 - Null hypothesis one was rejected ($p < .001$). Twenty-one percent of the variance in the post-test attitude scores of the subjects was accounted for by their membership in one of the treatment groups. This large difference provides a high degree of confidence in the effectiveness of the attitude change treatments.

Hypothesis 2 - Null hypothesis two cannot be rejected ($p = .29$). It can be inferred that the combined role playing experience is no more effective than only reading the persuasive communication.

Hypothesis 3 - Null hypothesis three cannot be rejected ($p = .84$). From the results of the test for H_2 and H_3 it can be inferred that there is no significant difference in the relative effectiveness of the three treatments.

The three null hypotheses were tested using pretest and post-test items. A post-post-test was administered. The magnitude of difference between the control group and the combined treatment group was smaller on the post-post-test than on the post-test comparisons. The treatment groups were still significantly different from the control group while not significantly different from each other.

Interpretations

The investigator concluded that attitudes toward science and elementary school science instruction can be modified by the use of an attitude-change technique with a persuasive communication as its basic mechanism.

Two factors were presented that should have contributed to groups RO and RRP being more influenced than group R. These factors concern the total amount of information to which each subject was exposed and the reiteration of the topics of the communication. A high level of acceptance of the ideas of the communication prior to the role playing may account for the failure of the treatment factors to create a difference between groups.

The results of the study do not support the value of role playing as a superior attitude-change technique to reading a persuasive communication.

Abstractor's Notes

The basic component of the study deals with attitudes of preservice teachers toward science and scientists. The total span of the study covered forty days. Since the results show only differences between control and combined treatment groups, the difference may not be the result of the specific treatment, but the result of any treatment. Perhaps preservice teachers were reacting to being in a treatment group regardless of the treatment. College students learn to give answers that they judge as appropriate to a specific situation. These students can present certain attitudes but not act in accordance with these stated attitudes.

Several possibilities exist to get at the development of attitudes! First, attitudes and changes in attitudes appear to be long-range phenomena. What effect might there be between the four groups in excess of ten days? What teaching behaviors might indicate that preservice teachers hold certain attitudes? Perhaps role playing may be a more suitable method to learn certain teaching strategies instead of attitude development. Was attitude development or rather modification of attitude the subject of this study?

The review of literature had little to do with role playing as a technique. It is not clear that the few studies cited dealt with science. No time parameter was given. If attitude change does occur, how lasting is this change? Even if the attitude modification does occur, what evidence indicates that affirmative action necessarily accompanies an attitude shift? Specifically, no studies were cited that even alluded to attitude changes of preservice teachers and concomitant teaching strategies that reflected these attitudes.

The sample was drawn from only four instructors. No control as to classroom events was cited. What occurred in these various classes during the forty days of the project was not discussed. The investigator stated that a setting was selected to make an experimental treatment related to science and science teaching seem a natural extension of the normal class objectives. Were these classes required or elective? If elective, these students may have already held more positive attitudes toward science and thus were easily subject to attitude modification by any of the three treatments.

Several unanswered questions remain concerning the attitude measure used as a pretest, post-test, and post-post-test. How was scoring completed? Students responded to one of seven points on a Likert scale. How were these responses converted to the scores reported such as a mean of 43.33 for group C on the post-test? No estimates of reliability or validity were reported. Since the ten science items were part of a 34-item instrument, was there an effect due to the 24 distractors? Sample items should be reported to give a clearer picture of the type of items used. Are ten items sufficient to gauge attitude modification of such a broad area as science, scientists, and scientific endeavors? When in the course of the academic year was the attitude instrument administered? If this study was completed at the beginning of the semester, there would be less chance of classroom-instruction variables contributing to the test results.

The persuasive communication was composed by the investigator. Was any pilot work done with the communication to identify reading level and the appropriateness of the 200 words per minute time limit? Perhaps a fifth experimental group who did not read either the persuasive communication or the innocuous communication would have been appropriate. The effect of reading any communication could have been explored.

The analysis of data does support the investigator's contention that attitudes toward science, as measured by a 10-item instrument can be modified. It does not appear possible to extrapolate as to the ultimate effect on the preservice teacher as measured by performance with children. Even if the attitude modification does occur, does it make any difference? If so, how do we know?

Evidently these students were not familiar with role playing strategies. The investigator reported that he observed very little new information or arguments from the role playing. If role playing was really the focus of the study as stated in the title, perhaps activities including role playing should have preceded the study.

The statement was made that the majority of the subjects were probably much in favor of the point of view of the communication before role playing. Possibly the role playing and the watching of the role playing did clarify the content of the communication. Could this same effect be created by a discussion among the subjects? The investigator does address himself extensively to possible factors that may account for the failure of the treatment to create a difference between groups.

The following statement was made that does not appear to be a valid generalization from this study as stated. "The implications from this study suggest that the explicit evaluation of the science programs in terms of the desired pupil activities and terminal objectives may be a better method of creating the desired science attitude in prospective teachers than indiscriminantly presenting science teaching materials, science curricula, and science teaching methods and hoping that the prospective teacher will synthesize this information in such a way as to develop a 'good' science attitude." The study as reported does not imply that persuasive communication and associated treatments will aid preservice teachers in becoming "better" teachers. Perhaps placing preservice teachers in public schools as part of their training would be more effective in attitude modification toward children than would a contrived role playing situation. Perhaps participation in a research project would accomplish an attitude shift concerning science and scientists. The unanswered question remains--Even with an attitude modification that is positive toward science, will the elementary teacher be more "effective?" It sounds reasonable to expect this outcome, but this particular study is not addressed to this issue.

If attitudes can be modified on a permanent basis; if the attitude measure is valid and reliable; if the sample was of sufficient size to generalize; if the persuasive communication was specific to the attitude modification desired; and if the attitude modification in fact makes a difference in the ultimate performance of the teacher, then the techniques reported may be an aid in teacher training.

EJ 042 226

Moon, Thomas C., "A Study of Verbal Behavior Patterns in Primary Grade Classrooms During Science Activities." Journal of Research in Science Teaching, Vol. 8, No. 2:171-177, 1971.
Descriptors--Educational Research, *Elementary School Science, Inservice Teacher Education, *Interaction Process Analysis, *Questioning Techniques, Teacher Education, Verbal Communication

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Lowell Bethel, University of Texas at Austin.

Purpose

The purpose of this study was to select and analyze verbal behavior patterns of both elementary school pupils and teachers while they were participating in classroom science activities.

The six hypotheses tested in the study are: There will be no significant differences in

1. the teachers' I/D ratios during science activities, before and after the introduction of SCIS teaching methods and materials.

2. the percentage of time teachers spend talking during science activities, before and after the introduction of SCIS teaching methods and materials.

3. the percentage of time students talk during science activities, before and after the introduction of SCIS teaching methods and materials.

4. the percentage of continuous student comment during science activities, before and after the introduction of SCIS teaching methods and materials.

5. the kinds of questions teachers ask children, before and after the introduction of SCIS teaching methods and materials.

6. the teachers' comprehension of the process aspects of science, before and after the introduction of SCIS teaching methods and materials.

Rationale

Many of the new elementary science programs require changes in the traditional role of the classroom teacher if they are to be successfully implemented. One method for determining the degree of role modification after introduction of a new elementary science program (SCIS) is to examine samples of verbal behavior patterns used during classroom science activities. Basic assumptions underlying this study are:

1. Verbal behavior selected for analysis is an adequate sample of the teachers' total classroom behavior.

2. The quantity and quality of teachers' talk largely determine student reactions.

3. The types of questions asked by teachers determine the quality of teaching and the level of thinking being stimulated.

4. The lessons observed are representative of the science lessons taught daily during the study.

The investigator does cite previous literature to which this investigation is related (i.e. Fischler and Anastasiow, Snyder).

Research Design and Procedure

The sample consisted of 32 primary grade teachers located in five Michigan school districts (DeWitt, East Lansing, Grand Hedge, Lainzsburg, and Williamston). Sixteen of the teachers used Science Curriculum Improvement Study (SCIS) methods and materials while the remaining 16 teachers used conventional science methods and activities. In addition, the 16 teachers using the SCIS methods and materials had been participants in a three-week SCIS workshop the previous summer during the investigation. This was the primary experimental variable.

Tape recordings of the verbal behavior of the 32 teachers were made during each science lesson observed from April 22, 1968, until March 27, 1969. Formal observation of the 16 SCIS-trained teachers began prior to their summer in-service workshop. The SCIS-trained teachers were observed five times and the conventional teachers were observed only two times. The recorded samples of verbal behavior were analyzed and evaluated using three instruments: The Flanders System of Interaction Analysis, the Science Teaching Observational Instrument, and the Science Process Test for Elementary School Teachers. The Process Test for Elementary School Teachers was a written test and was administered to the 16 SCIS-trained teachers prior to the participating in the summer workshop and again at the conclusion of the study.

The Flanders System of Interaction Analysis was used to obtain data to test the first four hypotheses. The Science Teaching Observational Instrument was used to evaluate the data to test the fifth hypothesis. The Science Process Test for Elementary School Teachers was used to evaluate the data to test the sixth hypothesis.

Findings

The findings of the investigation were as follows:

1. No significant differences were found between the two teacher groups in their I/D ratios, percentage of teacher talk, percentage of student talk, and percentage of continuous student comment during science activities.

2. No significant differences were found between the initial observation and final observation of those teachers employing the conventional science activities in regards to I/D ratios, percentage of teacher talk, percentage of student talk, and percentage of continuous student comment during science activities.

3. A significant difference was found in the I/D ratios of the 16 SCIS teachers as measured before and after the introduction of the new science methods and materials.

4. No significant differences were found in SCIS teachers' percentage of teacher talk, percentage of student talk, and percentage of continuous student comment during science activities.

5. A significant difference was found in the types of questions SCIS-trained teachers asked their students.

6. No significant differences were found in the SCIS-trained teachers' comprehension of the process aspects of science as measured by the Science Process Test for Elementary School Teachers.

Interpretations

The I/D ratios of the SCIS teachers did improve significantly. However, the researcher noted that the I/D ratio of the SCIS teachers decreased over time, indicating that they moved from a more indirect verbal behavior pattern approach to a more direct verbal approach during science activities. A possible reason offered for this decrease was the materials-centered approach of the science program. Directions given to pupils for working with the materials tended to depress the I/D ratios. An indication of this was the increase in the percentage of direction-giving (Flanders Category 6).

The SCIS teachers did significantly shift from asking low-order question types (requiring little cognitive skill) to higher order questions (requiring higher cognitive efforts).

Abstractor's Notes

Many of the new elementary school science programs emphasize inquiry skills including oral communication skills. The role of the teacher in providing a classroom environment which fosters the development of these skills is critical. Thus, change in the traditional role of the classroom teachers is required if these goals are to be realized. The investigator in this study attempted to assess the effect of in-service experience and new science materials in modifying the traditional role of the conventional teacher by evaluating samples of teachers' and pupils' verbal behavior. However, there appears to be a paucity of literature to support this view as cited in the investigation. Although space limitations considered, there should be more information given in this area. No summaries of the research literature are made. Thus, it is difficult for the reader to assess the importance of the investigation's outcomes.

A major shortcoming of the study is the description of the sample tested and the sampling procedures. No indication is given as to how the teacher samples were chosen and why. In addition, the teachers chosen are so poorly described that it might be difficult, if not impossible, to replicate this study because of this shortcoming. Such factors as race, age, years of teaching experience, and sex, as well as other sample traits are omitted. Further, the investigator does not describe the pupil sample as used in the investigation. There is no mention of age, grade-level, sex, or socioeconomic level. These variables are all critical to replication and are conspicuously absent.

Although the instrumentation used in the investigation is adequately described, no mention is made of reliability and validity coefficients. A reader unfamiliar with the instruments would be at a loss to determine their true value as used in this investigation.

Procedural bias may have been introduced by taping the control group only two times. Would there be a problem in randomly obtaining sufficient verbal behavior samples for analysis and evaluation? Perhaps this should have been considered. The investigator did not describe how he selected samples of verbal behavior for analysis and evaluation. How is the reader to know if there is any sampling bias here?

The results are presented in a concise manner. The written description is consistent with the data and inferences are kept to a minimum in the results. The analysis of data collected using the Flanders Interaction Analysis System does present some problems as the investigator admits. He does offer suggestions for procedures and methods to be used in future studies involving science classroom activities. Although the results of this investigation revealed that the verbal behavior patterns of teachers are altered after the introduction of SCIS teaching methods and materials, little or no change in the verbal behavior patterns of the children occurred.

EJ 079 319

Raymond, Anne, "The Acquisition of Nonverbal Behaviors by Preservice Science Teachers and Their Application During Student Teaching." Journal of Research in Science Teaching, Vol. 10, No. 1:13-24, 1973.

Descriptors--*Classroom Communication, *Educational Research, *Methods Research, Microteaching, Nonverbal Ability, *Preservice Education, *Science Education, Teaching Techniques, Verbal Ability

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Glenn H. Crumb, Western Kentucky University.

Purpose

Research conducted over the past several years has consistently identified silence or nonverbal teacher-student interaction as an area needing better description, analysis and study. There are, however, only a few studies reported that attempt to analyze the effect of purposefully designed teacher training activities upon nonverbal interaction in the classroom.

The study by Raymond investigated the effect of training preservice teachers enrolled in a methods course to identify and practice the skill of using nonverbal cues and silence. The parameters analyzed included time devoted to nonverbal behaviors, time devoted to congruent behaviors with students, number of positive nonverbal interactions and pupil perceptions of effectiveness of the student teachers who have completed the brief training program on nonverbal behavior.

Rationale

It is an underlying assumption of this and other research studies on "micro-criteria" that teachers who are able to exhibit each skill in a set of teaching skills (at or beyond the criterion level) will display greater overall teaching effectiveness than those teachers whose training does not concentrate upon the development of one skill at a time. In this research the use of nonverbal cues and silence during teacher-student classroom interaction was investigated to determine the effect of training of preservice teachers in using this skill. Although considered separate and apart from other teacher skills, the data collected and analyzed included both verbal and nonverbal interactions (classed as positive, negative and neutral) observed over the same period of time within the classes being taught by the experimental subjects.

Research Design and Procedure

The sample for the study consisted of twenty preservice secondary school science teachers, selected from a group of thirty-one who were enrolled in the micro-teaching experience during the fall and winter terms, 1970-71. The twenty subjects randomly assigned to the treatment and control groups were those preservice teachers who elected to complete their student teaching requirement during the winter or spring terms, 1971.

All subjects were exposed to the concept of micro-teaching via lecture and discussion conducted by the researcher at the initial meeting of a one-credit hour seminar. For each micro-lesson to be developed, the specific teaching skill (e.g. set induction) was identified, then a filmed model lesson was presented. Following this presentation, questions were answered. Then the subjects were asked to prepare a science lesson to focus upon the particular skill.

During the research phase, both the treatment and control groups developed and taught micro-lessons designed to develop the teaching skills of "set induction" and "using probing questions," according to a prepared schedule. The treatment group developed and taught a third micro-lesson designed to develop skill in use of nonverbal cues and silence. For each micro-lesson the procedure for developing the skill was the same; the skill was identified by lecture and discussion, a model was viewed on film, a second discussion/question period followed, micro-lessons were prepared and taught, micro-lesson critique by students and supervisor followed, then "reteach" sessions and subsequent critiques ensued until criterion level performance was reached. For the first micro-lesson, peer preservice teachers served as students and all sessions were videotaped. For the second and third micro-lessons, volunteer public junior high school students who were in study hall in the last two periods of the day served as students for the micro-lessons.

During their student teaching experience, following the training, each preservice teacher in the study was asked to select a class for videotaping. The tapes were used for analysis of teacher behaviors using the Biology Teacher Behavior Inventory (BTBI). An interobserver reliability of .81 was finally reached after initially finding agreement "too low to establish reliability."

The quality of the teacher-student nonverbal interaction was analyzed from classroom observation data collected by three cooperating teachers and the researcher using the % ALL. The % ALL was modified to allow coding of the initiating source of the interaction. Intercoder agreement of .79 on the Scott Index of Inter Coder Agreement was reached after training.

Data on student perception of teacher effectiveness were collected immediately following the videotaped lesson conducted by the student teacher. This was accomplished by having the members of the student teacher's class complete the Teacher Demonstrationg Rating Form.

The record of the videotaped teacher-student classroom interactions was encoded using the BTBI, according to the various forms of expressed behavior including "verbal," "nonverbal," "congruent," and "contradictory." The number of seconds of each behavior evidenced by the teacher was recorded and these data converted to percentage of total behavior observed during the class period. This procedure was used in order to allow comparisons of teacher data collected for class period of different lengths. Contradictory behavior, as a category, was subsequently eliminated due to low incidence.

Findings

The percentage of time that each teacher displayed nonverbal behavior was ranked across both groups and the Mann-Whitney U was applied to test the significance of difference in the amount of nonverbal behavior displayed. The hypothesis that "teachers who have identified and practiced the skill of nonverbal cues during a methods class will devote significantly more time to nonverbal behaviors during the student teaching experience" could not be rejected at the .10 level.

The percentage of time teachers spent expressing behaviors congruent with students was also calculated, ranked across groups and subjected to the Mann-Whitney U Test. The hypothesis that the treatment group would display significantly more time to congruent behavior with students was rejected.

The teacher-student interaction patterns resulting from the application of the % ALL was analyzed by determining the average number of observed incidences of teacher-initiated, positive, nonverbal interactions. These means were subjected to a t-test and found to be significantly different beyond the .05 level. The difference between the means favored the treatment group.

The scores of teacher effectiveness as indicated by the secondary school pupils in the classes taught by the student teachers were tabulated and the mean scores for the two groups subjected to a t-test for significance of difference in means. The calculated value of t was reported to be "not significant" at the .10 level and the hypothesis that the treatment group teacher would "be perceived as more effective" was rejected.

Beyond the results related to the hypotheses tested, the following findings were reported:

a. A significant positive correlation existed between the ranking of incidence of teacher initiated, nonverbal positive interaction and the rank order of teacher effectiveness (.05 level on Spearman Rank Correlation).

b. The control group devoted significantly more time to verbal interaction according to the BTBI classification. (1) "States knowledge"--verbal; (2) "States knowledge"--congruent; and (3) "Shows knowledge"--congruent. The treatment group spent significantly more time in "positive affectivity"--nonverbal (Mann-Whitney U).

c. Nonverbal teacher behavior for the treatment group made up 44 percent of their time while control group teachers devoted 34 percent of their time to nonverbal behavior.

Interpretations

The conclusion reached by the researcher was that "the experimental group of teachers differed significantly in their classroom behavior from the control group." The experimental group "devoted more time to nonverbal behaviors, including more teacher-initiated, positive nonverbal interactions with students in their classrooms," and had a larger total amount of time "spent in nonverbal positive affectivity." A further conclusion was that a causal relationship existed between the number of positive, nonverbal interactions which a student teacher initiates and one measure of teacher effectiveness.

In summary, the study indicated that inclusion of identification and practice in the skill of using nonverbal cues and silence in the preservice education of science teachers leads to more use of positive nonverbal interactions with students in the classroom during student teaching. From this finding, and the significant positive correlation reported between the number of positive nonverbal interactions and student perceived teacher effectiveness, there seems to be an implication that such preservice training improves teacher effectiveness and that such training should be included in science teacher education programs.

Abstractor's Notes

Because of the space limitations placed upon any author of an article for professional journal publication a reviewer frequently has a host of perhaps trivial, nagging, unanswered questions regarding research design, procedure and the collection and treatment of data. In most instances a review of the total research report and/or an interview with the author can retire most concerns or at least reduce them to an understandable "real world of operations" level. Nevertheless, questions concerning sample size and selection as well as treatment group versus control group experiences must be carefully scrutinized before generalizing too far afield from this reported study. The reported data collection and reduction procedures used in this study leave some unanswered questions about inferences based upon observations of behavior for the treatment and control groups. For example, was the exposure time to school students during the training period the same for the treatment and control groups? Did the control group develop and teach a "placebo" lesson to public school students to offset the treatment group's contract during the micro-lesson use of nonverbal cues and silence? What other experiences prior to student teaching may have been different due to scheduling treatment and control group membership separately?

It is observed that no mention was made in the report of the specific period of time over which the data were collected regarding teacher-student interactions using the _% ALL_. If sampling techniques were used, what were they and how were they determined? In addition, what method was used to determine how cooperating teachers and the researcher were assigned to collect data? Did they know which student teachers were in the treatment and control groups respectively?

If the above questions can be effectively neutralized, then it is appropriate to examine the data and its statistical treatment in search for "hard evidence" in support of rejecting or not rejecting the stated hypotheses as well as for implications and inferences for teacher education at large. Based upon the tabular presentations in the report, Table III, one is struck by the fact that on the average, the treatment group members displayed slightly less than five positive nonverbal teacher-initiated interactions and the control group teachers each averaged slightly over two. This is an exceedingly small number of incidences of observed behavior compared to the quantity of other behavior logged. For example, about 100 incidences of nonverbal behavior were recorded for each of the treatment group and control group. A "quick and dirty" analysis of the data in Table III leads to the observation that 70 percent of the control group teachers' initiated nonverbal behavior was neutral, while about forty percent of the treatment group teachers' initiated nonverbal behavior was neutral. It is also observed that about thirteen percent of the nonverbal behavior of the treatment group was negative whereas only four percent of the control group teachers' initiated nonverbal interactions fell into this category. This seems to suggest that whereas the report of the study focuses only upon nonverbal positive behavior, it may be advisable to examine the data in Table III for inferences regarding other types of nonverbal behavior as well. Such inspection might well include the number of incidences of behavior logged in both the verbal and nonverbal categories, as well as total incidences of behavior categorized by the cooperating teachers and the researchers. Although the data presented in Table I lead one to the conclusion that the treatment group devoted more time to nonverbal behaviors, the data in Table III indicate that the treatment group displayed about twice as many incidences of verbal behavior as did the control group while only a small difference was found between the two groups in the total number of incidences of teacher-initiated nonverbal behavior.

In addition to problems associated with the teacher-initiated behavior as presented in Tables I and III, there is the one of incidence of student-initiated interactions. The number of incidences of student-initiated interaction favored the treatment group (about 250 to 330). Most of this difference between the two frequencies of student-initiated behavior can be attributed to the large amount of student-initiated verbal behavior on the part of the members of the classes of the treatment group. Could this have some implications for the reported finding "that the number of positive, nonverbal teacher-initiated interactions correlates significantly with the student's perception of teacher effectiveness?" Could it be that the greater incidence of student-initiated verbal interaction implies something about the quality of teaching effectiveness as perceived by the students in the classes of the treatment group teachers?

The extended analysis of the behaviors of the two groups as categorized by the Biology Teacher Behavior Inventory does not provide sufficient information for reader analysis and evaluation. The statement that the computation of the U values "was hampered by the fact that 'tied data' interfered with the assignment of meaningful ranks to the teacher" is weak and indicates limited understanding of the statistical procedures for, and limitations and implications of, using the Mann-Whitney U as a statistic. Table VIII provides an interesting inference when coupled with the narrative. It would lead one to believe that the control group talked more, "States Knowledge," when the observations logged and reported in Table III indicate that the number of incidences of teacher-initiated verbal interactions are higher for the treatment group in the ratio of about 2:1.

In order to place the above comments into proper perspective it is necessary to examine the research scope and intent. Although not stated explicitly in the Journal, it is presumed that this study was conducted as part of a degree requirement. As such, it should be perceived to be a very fruitful undertaking. It presents some findings worthy of consideration and points out some quite obvious "holes" in research and practice in teacher training in general and science teacher training in particular. The shortcomings cited are meant to be flags for those who would accept, at face value, the results presented in the article. The limitations of the study are such as to deem such action unwise. Further research on this topic does seem warranted and this work of Raymond certainly provides a substantial contribution toward its initiation.

EJ 079 322

Renne, Thomas, Heidi Kass and Marshall A. Nay, "The Effect of Verbalizers on the Achievement of Non-Verbalizers in an Enquiring Classroom." Journal of Research in Science Teaching, Vol. 10, No. 2:113-124, 1973.
Descriptors--*Academic Achievement, *Behavioral Objectives, Discovery Learning, Educational Research, *Learning Processes, *Questioning Techniques, Science Education, *Student Characteristics, Verbal Communication

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Jal S. Parakh, Western Washington State College.

Purpose

The general purpose of this study was to obtain some knowledge of one type of classroom verbal interaction, namely questioning. Specifically, the following questions were investigated: (1) Do participants and non-participants differ in the extent to which they benefit, as measured by an achievement test, from the verbal queries of the participants? (2) What are some of the characteristics of students who tend to participate in asking questions, including sex, I.Q., sociometric rating, academiometric rating, and reading level, as opposed to those who do not? (3) What effect does guidance have on achievement, the degree of participation, and the nature, quantity, and quality of the questions asked?

Rationale

The desirability for students to learn science by asking questions, holding discussions, and interacting with each other is often mentioned in the science education literature, but there is a shortage of specific knowledge as to how effectively the students learn from one another.

While classroom enquiry and verbal interaction entail more than just question asking, the study was limited to questioning. The basic assumption is that questioning is an important aspect of classroom enquiry. More specifically, the Suchman questioning technique was adopted for this study because it enables a certain degree of experimental classroom control and involves student-initiated questioning, which is generally considered desirable. The intent of this investigation was not to study the process of enquiry as defined by Suchman but some aspects of the questioning mode of verbal interaction.

Suchman found that children who received enquiry training became more fluent in questioning, asked more analytical questions and fewer abstract conceptual questions, particularly of the diffuse type, than did the control group.

However, when only some of the students ask questions, to what extent do the remainder benefit and to what extent are the verbal probes of the questioners taken as cues by the others and how is science achievement influenced? Gallagher found significant differences favoring expressive boys over non-expressive boys on both a teacher-made high school biology test and a BSCS Test. The same trend was noted for girls but the difference was not statistically significant. Boys contributed more to verbal activity than girls and expressive students were found to have a higher mean aptitude score than non-expressive students.

The amount of direction of the enquiry process desirable for concept development has been investigated among others by Craig, Kittle, Butts, Scandura, Butts and Jones, Tanner and Thomas and Snider. (No statement of findings of these studies was reported in the article.)

Research Design and Procedure

Twelve randomly selected grade VIII science classes were in turn assigned at random to three treatment groups of four classes each. The initial supply of information was the same for all classes in all treatment groups and was provided by one silent 8mm film loop (unnamed) in the Science Research Associates Inquiry Development Program series.

Treatment A, or the guided enquiry treatment, consisted of the film, a test based on the film content which was called a pretest and was also considered to act as a guide for the subsequent enquiry session, and then the same test was readministered after the enquiry session and called a posttest (emphasis added).

Treatment B, or the unguided enquiry treatment, was not to administer the pretest or guide, but to engage in the enquiry session immediately after the film was viewed and then to take the same posttest as Treatment A.

Treatment C, or the self-guided enquiry treatment, consisted of the film followed by the students being asked to submit five written questions answerable by "yes" or "no," then the enquiry session, and the same posttest as in A. The written questions were also used to identify the enquiring and non-enquiring non-participants.

Three enquiry sessions based on the same three (unnamed) films were conducted with each class by the principal investigator in three subsequent meetings. A sample enquiry session was conducted with each class prior to the experimental treatments. During the first three meetings with each of the classes, the following were administered: the Lorge Thorndike Non-Verbal Test, Level 4, Form A; the Gates Reading Survey, Form 3; a sociogram; and an academiogram. In addition, an estimate of the extent of questioning in which each student engaged under normal classroom conditions was obtained by asking each teacher to rate his students on a five point scale, ranging from one for a person who rarely asked questions to five for a person who asked many questions.

The three achievement tests were based on the content shown in the three films. Each of the three ten-item tests had the four-option multiple choice format. Of the 30 items, 14 items were judged to be at the two lower levels of Bloom's Taxonomy (knowledge and comprehension) and the remaining 16 items were judged to be at the four higher levels of Bloom's Taxonomy. Three achievement scores, namely, total achievement, achievement on the two lower levels and achievement based on the four higher levels of Bloom's Taxonomy, were separately used in the analysis. The KR-20 reliability of the composite achievement measure was 0.76.

Findings

The major findings are listed below.

a. Significant Spearman rank order correlation coefficients were found between the classroom teacher's rating of students' verbal participation and the number of questions asked in the enquiry session in 11 of the 12 classes.

b. In all treatment groups the mean I.Q. of the participants (students who asked at least one question per enquiry session or a minimum of three questions) exceeded that of non-participants (students who asked less than three questions). With I.Q. controlled, the mean achievement of participants was higher than that of non-participants but significant differences were found only at the four higher level categories of Bloom's Taxonomy.

c. With adjustments for ability, the highest mean achievement in both the composite and higher level categories was attained in the group that was given the pretest and called the guided enquiry group or Treatment A.

d. Sex and participation were found to be dependent upon each other with more boys and fewer girls participating. The Chi Square test was used.

e. Participation was found to be independent of sociometric position in the classroom but not of academiometric position.

f. The number, proportion and cumulative proportion of questions were also classified according to the Suchman categories. The Kolmogorov-Smirnov two-sample test was used to test for significant differences in distribution along the categories. At the 95 percent level of confidence, the question distributions were significantly different between treatments A and B and between B and C but not between A and C. It was not possible to make statistical adjustments for the differences in I.Q. since the frequency distribution for fluency of verbal enquiry departed greatly from normality. Of all the questions asked by each group, Verification questions accounted for 54 percent in treatments A and C and 60 percent in treatment B. In terms of Implication questions, the sub-categories of diffuse questions accounted for 5.2 percent, 8.9 percent and 4.5 percent for treatments A, B, and C and the sub-category of directed questions accounted for 19.8 percent, 8.1 percent and 17.5 percent for treatment A, B and C respectively. Both groups A and C asked many more specific cause and effect questions than did group B.

Interpretations

From the data in the study it appeared that student verbal participation in the classroom was associated with achievement at the higher cognitive levels of Bloom's Taxonomy.

Guidance appeared to have little effect on achievement at the lower levels of Bloom's Taxonomy. Guidance did appear to result in better achievement on the higher levels and the composite score.

Participants tended more often to be boys. This may be due to social reasons or to a less adequate content background.

Participants had higher I.Q.s than non-participants.

Sociometric choices did not appear to be based upon academic achievement in the classroom.

Abstractor's Notes

After reading the article the abstractor was hard-pressed to find a clear statement of just what effect(s) verbalizers had on the achievement of non-verbalizers in an enquiring classroom. If the major purpose of the study, as implied by the title of the paper and portions of the statement of the problem by the authors, was to determine the effects of participants on non-participants then that purpose seems not to have been achieved.

A second purpose of the study was to determine some of the characteristics associated with participation. The considerable amount of data collected on the students' characteristics such as their sex, I.Q., ability level, sociometric and academiometric position, etc., and subsequent analysis must indeed have required considerable effort and time. The findings were probably those that many, if not most, experienced classroom teachers would have predicted and thus the study served the important function of providing research evidence for expectations of experienced teachers (or at least the abstractor's assumption about predictions of experienced teachers).

A third purpose of the study was to investigate the effect of what the authors termed "guidance" on achievement as well as degree and nature of participation. In the article about half a dozen studies are simply listed as investigations related to the amount of direction provided in the enquiry process. Thus one is unable to assess from the article whether "guidance" as used by the authors was the same as that used by the other investigators or something similar or significantly different. Guidance as described by the authors was of two types: (1) in treatment A, a pretest was administered after the film but before the enquiry session and (2) the students in treatment C were asked to submit five written questions after seeing the film but before the enquiry session.

Some questions remain in the mind of the abstractor. Since the same test was used as a posttest, how much of the gain in achievement was due to taking the pretest and how much due to the participation in the enquiry sessions? To what extent were the questions of students in treatment A directed to finding the answers to the questions on the pretest?

The above comments and questions refer primarily to the statement of the problems and the purposes of the study. A number of questions occur regarding the procedures used. In the article it is stated that twelve grade VIII science classes were randomly selected. What was the size of the population from which the sample was selected? The enquiry sessions were based on three 8mm film loops in the SRA Inquiry Development Program. What were the names of the films? What was the nature of the pilot test that was used to determine the suitability of the films? What are the specific objectives of the films shown? Are the films suited primarily to develop enquiry skills as defined by Suchman and secondarily to develop concepts through enquiry?

In raising the questions and making the comments listed above, the abstractor recognizes that constraints upon the authors in terms of the length of the article may have been responsible for the absence of some or all of the information that would have constituted a more complete report of the study. Thus, the questions and comments should not be misconstrued. It seems clear to the abstractor that considerable effort was expended on the study and interesting findings have been reported.

EJ 037 867

Riban, David M., "On the Ability to Infer Deficiency in Mathematics From Performance in Physics Using Hierarchies." Journal of Research in Science Teaching, Vol. 8, No. 1:67-82, August, 1971. Descriptors--*Evaluation, Instruction, Mathematics, *Physics, *Programed Instruction, *Remedial Instruction, Secondary School Science

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Robert L. Steiner, The Ohio State University.

Purpose

The purpose of the study was twofold:

1) to develop and evaluate a system to infer deficiencies in requisite mathematics abilities of students enrolled in high school physics, and 2) to investigate the effect of remedial instruction in requisite mathematics abilities on physics achievement.

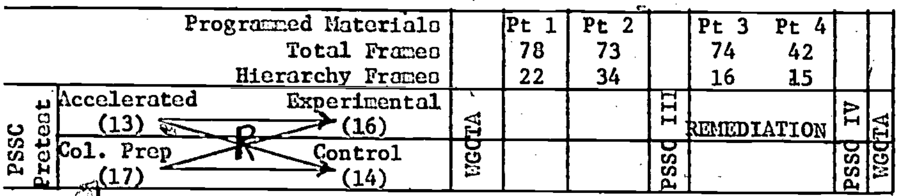
Rationale

The absence of certain mathematics abilities has been shown to adversely affect learning in science (Redfield and Atwood, 1966); therefore, poor student achievement in physics may be based in part on unlearned requisite mathematics abilities. Identification of hierarchies and individual diagnosis of requisite abilities has been successful in experimental work; however, to be practical for common use, some statistical diagnosis of absent hierarchical abilities for normal classroom work will have to be developed. If a number of physics problems are analyzed and hierarchies of mathematics abilities necessary for successful completion of each problem are prepared, it should be possible to infer student deficiencies in mathematics abilities from individual performance patterns of correct and incorrect solutions to the problems.

Research Design and Procedure

An accelerated 11th grade physics class (n = 13) and a 12th grade college preparatory physics class (n = 17) were used in the study. Students in each class were randomly assigned either to experimental (n = 16) or control (n = 14) groups. Both groups received a 10 week programmed instruction unit on light. The unit was divided into 4 sections and consisted of 267 frames, 87 of which were designated as hierarchy frames. In the 87 hierarchy frames, 163 requisite mathematics abilities were identified and remedial sequences were developed for 42 of the abilities occurring most frequently. The experimental group received remedial mathematics instruction during sections 3 and 4 of the programmed materials. The decision to remediate a student was made on error rates on each of the 42 mathematics abilities.

Student achievement in physics was measured by PSSC course tests. Two PSSC tests had been administered to the classes at appropriate intervals prior to the start of the 10 weeks of programmed materials. PSSC Tests III and IV were administered after sections 2 and 4, respectively. Also, critical thinking ability was measured using the Watson Glaser Critical Thinking Appraisal (WGCTA). The design of the study is illustrated in the following figure.



Findings

Physics achievement for both groups as measured by PSSC Tests III and IV was found to be significantly higher for the materials presented using the programmed instruction format than for the prior instructional method. The experimental treatment (remedial mathematics instruction) did not produce any significant differences in the physics achievement of the experimental and control groups as measured by PSSC Test IV. The gain on the WGCTA was significant.

No significant differences between the experimental and control groups in physics achievement of PSSC Test IV were found when covariates including the WGCTA pretest scores, PSSC Test III scores, WGCTA gain scores and number of physics program cards completed were used.

Fifty-six percent of the 82 remediations prescribed during the study were judged as misdiagnosed; therefore an ex post facto examination of the diagnosis procedure was undertaken. Performance patterns represented below were developed for each student on each of 52 mathematics abilities identified for part 1 and 2 of the materials. A plus was recorded in each of the quadrants where the particular performance indicated was exhibited at least once. All sixteen of the possible arrangements of + and - were exhibited in the 1456 individual patterns (28 students x 52 mathematics abilities). The 16 possible patterns were classified in the following manner:

	Part 1	Part 2
Correct		
Incorrect		

1) no change (2 patterns) consistently correct or incorrect responses on parts 1 and 2.

2) inconsistent with valid measure (4 patterns) incorrectly responding on part 2 after correct responses on part 1.

3) consistent with valid measure, incidental learning (2 patterns) responding correctly on part 2 after incorrect responses on part 1.

4) no data (8 patterns) no clear information on student performance trends.

Using all student patterns, the observed frequency of + and - for each quadrant was calculated and the probability of each of the 16 patterns calculated. A chi-square test of predicted and observed numbers was made for each of the 4 categories of performance patterns. The test was significant at $p < .001$, indicating that the patterns were not simply a collection of random data. Next, 28 of the 52 mathematics abilities were removed from consideration because of unequal or limited distribution in parts 1 and 2. Performance patterns for the remaining abilities were calculated and a chi-square test of predicted and observed numbers was made. The result was again significant, indicating that the performance patterns could be used to measure deficiencies in mathematics abilities. The performance pattern method was improved even more when only 11 mathematics abilities with maximum independence were used. Had performance patterns been used for making remediation decisions, instead of error rates as used in the study, incorrect remediations would have been reduced from 56 to 35 percent.

Interpretations

Failure of the experiment to improve physics achievement by remediation in mathematics abilities may have been related to an ineffective system of diagnosing the need for remediation. Perhaps diagnosis based on a tally system and performance patterns would have yielded different results.

Two performance patterns of consistent measure but indicative of incidental learning were much higher than anticipated and incidental learning was four times as important a source of mathematics remediation as the procedure undertaken in the study. If incidental learning is common in physics instruction it might indicate that specific mathematics deficiencies are not a serious problem. Alternatively the physics programmed materials may have helped students learn remedial mathematics and this is why all groups did well.

Abstractor's Notes

Lack of mathematical abilities is cited as a hinderance to success in physics by both students and teachers. This study addressed itself to this problem.

Although the stated purpose was twofold, the experimental portion was limited by the ineffective diagnostic system for identification of those in need of remediation. One is left with the impression that the results of remediation of requisite mathematics abilities on physics achievement were insignificant and as a result, rather than concluding that the treatment was ineffective, an alternative explanation was offered. It was suggested that the method of diagnosis of need for remediation may not have been effective and that a more careful analysis of the method of identification was in order. Since the diagnosis for remediation was of major importance to the study, it would seem that this should have been done and evaluated before the study was begun, not ex post facto.

The title of the study and the early discussion of hierarchies implies a greater dependence on Gagné type hierarchical structure than is warranted. The author was unable to develop agreeded upon hierarchies of requisite abilities. Instead, a number of mathematical abilities, which may or may not have been dependent on each other, were identified for each frame. Albeit the author did acknowledge this problem, indicating "... each 'hierarchy' was judged as an unordered set of requisite abilities for the problem in question." The inclusion of hierarchical structure could have been omitted and, instead, the problem approached in terms of a number of separate mathematical abilities.

It was concluded that the programmed instructional format was more effective than the prior instructional format because of significantly higher scores on the PSSC Tests III and IV. Both the experimental and control groups used this mode, hence there was no control. Alternate hypotheses could be offered for the significantly higher scores on PSSC Tests III and IV.

Although significance tests were made between the experimental and control groups, most of the data was presented in terms of either the initial classes or the four subgroups. It would have been helpful and perhaps more meaningful to have the data also presented in terms of experimental and control groups. In addition, initial differences were indicated between the accelerated and college preparatory classes, but the article does not indicate if there were any initial differences between the experimental and control groups. Even though random assignment of the students in both classes to experimental and control groups was made, significant differences between the two groups could have existed and should have been tested or reported if a comparison was made.

The raw scores on the PSSC tests of the two classes and also for the regional student data used for the standardization of the scores would have been helpful. It is somewhat surprising that the accelerated and college preparatory classes both scored below the regional norms. The article does not give any means to verify the results presented.

PSSC Test IV used to measure physics achievement may not have been a valid measure of the effect of remedial mathematics instruction. No analysis was made or indicated to determine if the mathematical abilities remediated were necessary for successful completion of the questions on the test. Perhaps all abilities remediated were not needed in the PSSC Test IV and hence would not affect physics achievement.

The remediation sequence included an entry problem requiring the mathematical ability in question. If the student could do the entry problem he was directed to an exit problem also requiring the ability.

If he could correctly work the exit problem he did not work through the sequence. The study indicated that 56 percent or 46 of the 82 remediations were judged as misdiagnosed. This suggests that only 36 remediation sequences were done for parts 3 and 4. This is only a little more than 5 percent of the possible 672 remediations (42 mathematical abilities x 16 students in the experimental group) for parts 3 and 4. It could be that all 42 mathematics abilities were not found in parts 3 and 4; therefore the total number of possible remediations would be smaller. In any case it seems as though so few students were initially lacking the mathematics abilities that the effect of remedial instruction had little potential to have an effect.

Some results presented in the study do not seem possible from the data. In the "Results" section (pages 71-72) two of the mean differences given do not agree with the standardized data given in Table II. The average gain between PSSC pretests and PSSC Test III is given as 6.63 points with a gain of 4.00 points for the accelerated class and 8.65 points for the college preparatory class, but the data in Table II indicate an overall mean gain of 4.87 points with the accelerated class gaining 1.81 points and the college preparatory class gaining 7.21 points. Also in the "Results" section, the discussion indicates 1201 hierarchy frames contributed data, but Table I indicates that the total number of hierarchy frames for both groups totals only 1021. The cards per individual for the accelerated class reported in Table I also seems in error. It appears that the author divided the number of remediation cards by all the students in the accelerated class, whereas only 7 of the 13 students were in the experimental group and hence participated in the remedial frames.

Seven of the nine (no data) performance patterns presented on page 74 are theoretically possible, but, if they occurred with any great frequency, it would suggest that the mathematics were not very evenly distributed throughout the program and the mathematics abilities necessary for part 1 were not necessary for part 2 and vice versa. If this were true throughout the four units, it would not be surprising that remediation based on one unit would not be of help for the next.

On page 75 the number of students used for calculation switches from 30 to 28 for no apparent reason and without explanation.

Calculations based on data in Table IV (page 78) give chi-square contributions of 37 and 86 respectively for the "No Change" and "Inconsistent" categories and not 45.1 and 82.9 as reported. Similarly in Table VI calculations based on the data give a chi-square contribution of 323.7 for the "Consistent" category rather than the 347 indicated in the table.

The discussion on page 76 considered an examination of the more evenly distributed mathematics abilities over parts 1 and 2. Twenty-eight of the original 52 abilities were removed from analysis, leaving 24 - not 26 - as indicated in the discussion and in Tables V and VII. The use of 26 instead of 24 also causes an error in the number of performance patterns which should be 673 and not 728 as reported.

The use of 728 probably produces errors in the frequency presented in Table V (page 76).

In addition to errors, there were several confusing points in the article. Some of the results are not clear from the data and discussion presented, all of which contribute to the difficulty in reading and interpreting the article. Perhaps additional discussion would have clarified some of the confusing points and apparent errors in the article.

EJ 079 320

Riechard, Donald E., "Life-Science Concept Development Among Beginning Kindergarten Children From Three Different Community Settings." Journal of Research in Science Teaching, Vol. 10, No. 1:39-50, 1973.
Descriptors--Cultural Factors, *Early Childhood Education, *Educational Research, *Kindergarten Children, Performance Factors, *Science Education, Social Factors, *Socioeconomic Background

Expanded Abstract and Analysis Prepared Especially for I.S.E. by David H. Ost, California State College.

Purpose

The basic question investigated in this study was: Do children from three different community settings (inner-urban, outer-urban, and rural farm) vary significantly in the ability to conceptualize selected aspects of life science at the time they begin kindergarten? The study also addressed the parallel problem of identifying those factors which are useful predictors of science concept development among beginning kindergarten children.

Rationale

A major premise of the study was that "concepts" are the basis or structure for problem solving and other higher order thinking processes. If this is true then the assessment of the ability of children to conceptualize science or the concepts with which children enter kindergarten is important to science curricula and instruction. (It is not entirely clear which of these is really being investigated.) References are made to the increased attention being paid to early childhood education and the resulting associated research. This information is used to amplify the importance of knowing the relationship of a child's background to his/her conceptualizing ability. It is pointed out that little research has been done in this area particularly as related to science education.

Research Design and Procedure

The investigation was of a descriptive nature; as such, no cause and effect relationships can be clearly defined. Six phases of the study are reported.

The first phase consisted of the development and pilot testing of the concept assessment instrument, Life-Science Concept Acquisition Test (L-SCAT) for children ages 4-6. The L-SCAT consists of 21 sets of colored pictures and an accompanying interview schedule. The pictures represent 35 concept items representing the seven BSCS content themes. A panel of judges critiqued the L-SCAT.

The instrument is an individually administered, picture-stimulus structured interview. The subject responds nonverbally and verbally about the pictures. Questions by the interviewer are based upon subject responses. Scores are attained for the nonverbal portion, the verbal portion (the tape record is scored in accordance with a criterion scale), and on each conceptual scheme. The instrument is not published in the report.

The pilot test was made with two populations: kindergarteners of Caucasian blue collar workers (n = 18) and pre-kindergarten children enrolled in a Head Start program (n = 13). No discussion is provided as to why these populations were selected. Item analysis, teacher ratings of children, as well as test-retest procedures were utilized to determine the validity and reliability of the instrument. No reliability coefficient was given.

The second phase of the reported investigation was the selection of the population. The three community settings were defined on the basis of: geography; percent of families receiving Aid to Families of Dependent Children; occupation of residents; and, distribution of inhabitants. Additional socio-cultural data concerning the child's background and experience were taken from a Kindergarten Enrollment Information Form.

Phase three was an attempt by the investigator to build a rapport with the subjects before collecting data from the subjects through the administration of various instruments. The administration of L-SCAT to 17 children from each community setting occurred in phase four. The subjects were selected at random by means of a table of random numbers. Phase five of the study was the administrating of the California Short-Form Test of Mental Maturity, Level 0, 1963 Revision. Phase six was the analysis of data. One way analysis of variance was employed to examine differences among scores made on the L-SCAT by subjects from the three community settings. This was followed by the use of Tukey (a) procedure as a means of determining the source of significance. Step-wise multiple regression analysis provided information concerning the value of predictors of performance on the L-SCAT.

Findings

Beginning kindergarten children from the three community settings differed significantly in performance on all the L-SCAT measures (nonverbal, verbal and total). The order from highest to lowest being "outer-urban," "rural-farm," and "inner-urban." Analysis of the audiotapes suggested that much of the variance may have been in fact due to differences in concepts and concept formation unrelated to science. The investigator listed examples of such concepts as "under," "before," "different," etc.

The step-wise regression analyses indicate that I.Q. scores were the most useful predictor of L-SCAT performance. Chronological age was the second variable entered into the equation. When only socio-cultural variables were used in the regression analyses for total score on the L-SCAT, the subject's mothers education was the first variable entered (most useful predictor). Younger siblings, "extended family" and zoo visits were entered respectively with increased power of prediction. The "only child" out-performed the child with siblings. The "extended family child" did not perform as well as the child who did not live in an extended family environment.

Interpretations

The investigator points to the ability of children to deal with the abstract concepts of time, change and relationships as important attributes related to some science concepts such as evolution. This interpretation is based, in part, on the results of subject's performance on the evolution theme (which is apparently one of the BSCS themes included in the L-SCAT). This supports the premise that science education for young children should not be drawn strictly from the domain of science; physical, mental and sociocultural factors must be given more than superficial consideration when designing science education programs.

Chronological age is cited as being a factor to which more attention needs to be paid. The range of 7.8 months in the chronological age of the subject represents 10 percent of the total time an average kindergarten child has lived. This fact is particularly relevant, according to the investigator, in that it is cited that 50 percent of the "intelligence" an individual has at age 17 is supposedly developed by the time the individual enters school.

Attention is drawn in the report of the study to variations in L-SCAT performances of children as a function of "birth-order rank" and other related factors. The investigator raises the possibility of aiming early childhood science education at the mothers of young children. This notion is amplified further by the suggestion that since children already vary significantly in development prior to formal school it may be appropriate to affect "other people" in the child's environment who would be able to influence the development of science and other concepts.

Abstractor's Notes

More questions are raised as a result of this study than are answered; that seems to be the nature of a descriptive investigation.* Descriptive investigations may have greater latitude in the design, greater leeway in the foci of study and produce results which are less elucidatory. However, the investigator has other constraints and responsibilities which are not generally associated with

---

* A descriptive investigation, as considered by this abstractor, is one which is more decision-oriented than conclusion-oriented; describes rather than explains; suggests the existence of relationships rather than demonstrating cause-effect; or, it may contribute to the solution of a specific or practical problem rather than explain the basis of the problem.

traditionally designed experimental investigations. To have its greatest use, a descriptive study must provide as much detail as feasible. In this study the investigator neglected to provide data concerning the reliability of the instrument; an instrument upon which the entire investigation was built. No rationale is provided for the populations utilized in the pilot testing of the L-SCAT. Why was the instrument not field-tested with the population which subsequently scored the highest?

The investigator provides no rationale for the use of the F-test with a relatively small population. It appears that the choice was based on the power of the test. In a similar manner, the establishment of the alpha level (.05) appears to have been made to increase the power of the test. For what reason? If calculated ratios are being utilized as descriptors, then the level of significance or the power of the test is not all that important (except for the reviewers of articles submitted for publication). The importance of significant differences in descriptive studies of the type reported here is of questionable value. It would have perhaps been more prudent and proper to discuss at greater length the Type 1 and Type 2 errors associated with the study. The use of the F-ratio in conjunction with such discussions is of greater meaning in descriptive analysis than simply knowing whether "it is statistically significant at the .05 level."

Similarly, it would seem that a discussion of the rationale behind the choice of the Tukey (a) procedure over the Scheffe method would be of value. The latter is considerably more conservative with respect to Type 1 errors than is the Tukey (a) method. Such information is important if the investigator elects to report his/her investigation in a descriptive fashion, let alone design it as such.

Step-wise regression analysis is a very useful tool in descriptive studies. Again, the value of such a statistical tool may not lie as much in the statistical significance of entered variables as in the order of entry. The list of independent variables compiled in the study is lacking. Hence, much descriptive information is lost. It would be of value to know which factors were found to be of little or no use in predicting L-SCAT scores. It would also be helpful to know how the independent variables were selected. (Given a list of 50 random variables, chances are that several will be "statistically significant" at the .05 level.) Of course, those researchers interested only in significant differences using different statistical tests are not impressed with such information.

Although the investigator indicates that the subject receives a score on each conceptual scheme found in the L-SCAT nowhere in the report is there a discussion of this item. (No significant differences?) To the researcher in science education these data would seem to be as useful as, if not more important than, the verbal and nonverbal scores.

A comment in passing has to do with the significant figures used in this study and the majority of educational research. It is difficult to ascertain the reason behind calculating a mean score (with an n = 17) to four decimal places. The science concept of significant figures seems to be undeveloped among education researchers.

In summary, the study may be timely although the report is deficient in some aspects. With the increase in attention being paid to research in early childhood education the results of studies such as these may pay large dividends. Multiple discipline studies may provide much greater insight into the areas of early childhood science education. It is hoped that investigators who design studies which are intended to be evaluatory or descriptive report their results in a manner congruent with that philosophy. No doubt a full report of the investigator would answer many of what some would consider mechanical or diminutive questions raised in this abstract. Unfortunately the reader has nothing but that article from the Journal of Research in Science Teaching which provides: less than adequate information about the sociocultural, physical, and mental variables used in the study; questionable background and general information concerning the major instrument in the study (L-SCAT); little or no information specific to science education or instruction; and, little rationale for much of the design of the investigation.

EJ 049 190

Siegel, Betsy Davidson and Ronald Raven, "The Effects of Manipulation on the Acquisition of the Compensatory Concepts of Speed, Force, and Work." Journal of Research in Science Teaching, Vol. 8, No. 4:373-378, 1971.
Descriptors--*Cognitive Development, Concept Formation, *Elementary School Students, Force, *Instruction, *Learning Theories, Psychology

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Mary Budd Rowe, University of Florida.

Purpose

Do students who actually manipulate materials (variables) learn concepts of speed, force and work as well or better than do those who see the concepts demonstrated?

Rationale

Most modern science programs stress the value to the student of actually performing activities and conducting investigations of various magnitudes. This activity-based approach to science instruction is meant to help students acquire concepts, learn to control and manipulate variables, and to develop some sense of the tentative nature of knowledge

Huttenlocher (1), however, showed that students who watched the experimenter do the manipulations rather than performing manipulations themselves, performed better on problem solving tasks than did the students who actually worked with the materials. She suggested that manipulation by the students may have interfered with the processes of remembering and interpreting information.

Siegel and Raven challenged Huttenlocher's findings which were at the time already nine years old. They argued that research by Inhelder and Piaget showed that children in the middle stage of concrete operations can solve problems which involve compensating variables by manipulating one variable at a time and comparing differences in the other variable. Nine to ten year olds must be presented, they reasoned, with concrete objects. If only verbal data or viewing in a demonstration is allowed, children cannot set up the problem in such a way as to make what happens meaningful.

Research Design and Procedure

To find out under what conditions fourth graders were more likely to be able to coordinate two or more variables that reciprocally oppose each other, Siegel and Raven assigned each of 120 fourth graders to one of three treatment groups. The manipulation and demonstration groups received nine hours each of instruction. Each child in these two categories received a workbook prepared by the investigators.

This workbook contained schematics for each experiment and questions to be answered. Each child in the manipulation condition had his own set of equipment. Children in the demonstration group watched experiments being done. At the conclusion of instruction each child was tested individually. Apparently there was no placebo for the control group.

Three contrasts were made. First, 20 students were selected at random from each group to determine whether the three groups were in fact equivalent on the outcome measure (Kruskal-Wallis test). They found that at least one linear contrast was different from zero so the next problem was to find out where the difference was. A procedure described by Marascuillo and McSweeney showed that the significant difference was between the control group and the treatment groups. But, in the third step, the linear contrast between the treatment group means was not significant, i.e., the demonstration and experiment treatments were equally effective.

The possibility still had to be considered that one or more of the concepts, speed, force, or work was better acquired under one treatment condition than another. The Friedman two-way analysis of variance performed on the remaining 20 students in each group yielded no differences among the task categories.

Findings

There is no difference between the manipulation and demonstration treatment conditions on the outcome measure. Students learn equally well concepts of speed, force, and work under either training condition. The trained students do better than the untrained students.

Interpretations

Siegel and Raven maintain that their work shows that training can stabilize a compensation scheme in fourth grade children.

Abstractor's Notes

The financial implications of this study are enormous. The investigators show that training by demonstration is as effective as training by active laboratory involvement, at least so far as acquisition of compensation schemes goes. It would seem that Huttenlocher's forecast had merit -- although oddly enough -- Siegel and Raven never return to a discussion of their findings in the context of Huttenlocher's work which provided the impetus for the study. Does manipulation interfere with the tasks of remembering and interpreting, as Huttenlocher suggested, or are the opportunities for rehearsal of ideas less likely to accompany a laboratory activity? Dewey often commented that activity without reflection is a waste of time.

It may be that a more refined question needs to be asked before we stop buying equipment in class lots and begin returning to a demonstration mode. Perhaps some students do well under one kind of instruction, e.g., manipulation and others do well under another, e.g., demonstration. A post hoc analysis would quickly show whether such were the case. In short there could be an aptitude-treatment interaction. We might surmise, for example, that the higher ability segment would do better under the demonstration condition while the lower ability segment would thrive under the manipulation condition. This kind of result would fall in line with recent studies in elementary science which show that low SES groups appear to benefit from laboratory type programs in the primary grades while middle and upper SES groups do not (2).

The control group received nothing but the outcome test, not even training in the vocabulary which was central to the outcome measure. Thus, on the one hand, the finding that the instructed groups do better is in one respect trivial. On the other hand, the fact that students can learn a compensatory scheme through specific instruction is not trivial. In so far as an evaluation of the Huttenlocher proposition goes we have in this study neither infirming nor confirming data since the two treatment groups do not differ significantly. However, the mean score for the demonstration group, while not significant, is higher on each of the three concepts, and one wonders what pattern would emerge with a larger sample. If Siegel and Raven still have their original data, it would be nice to get some preliminary idea as to how tenable the aptitude-treatment interaction hypotheses is. Macbeth, for example, showed that kindergarten children who manipulate science materials attain science process skills better than those who do not have this opportunity. At the third grade, however, the manipulation group did no better than the non-manipulation group (3).

The outcome measure appears to be closely related to the training content. So we do not know whether a generalized compensatory scheme was learned or whether compensation of three variables, once acquired, facilitates acquisition in another context. Neither do we know whether retention for the two experimental conditions is equally good. It was not necessarily the job of the investigators to answer these questions --but someone should, sometime.

References

1. Huttenlocher, J., "Effects of Manipulation of Attributes on Efficiency of Concept Formation." Psychological Reports, 10:503-509, 1962.

2. Rowe, Mary Budd and L. Deture, A Summary of Research in Science Education, 1973. New York: John Wiley and Sons, 1975.

3. Macbeth, Douglas Russell, "The Extent to Which Pupils Manipulate Materials and Attainment of Process Skills in Elementary School Science." An Abstract of a Thesis in Secondary Education, Pennsylvania State University, September, 1971.

EJ 079 839

Summerlin, Lee and Marjorie Gardner, "A Study of Tutorial-Type Computer Assisted Instruction in High School Chemistry." Journal of Research in Science Teaching, Vol. 10, No. 1:75-82, 1973.
Descriptors--*Chemistry, *College Science, *Computer Assisted Instruction, Educational Research, *Instruction, Instructional Media, *Science Education, Secondary School Science, Teaching Procedures

Expanded Abstract and Analysis Prepared Especially for I.S.E. by J. Dudley Herron, Purdue University.

Purpose

This research compares the effectiveness of a tutorial (as contrasted with drill-and-practice) CAI program in high school chemistry with conventional instruction.

Rationale

Although a number of CAI programs have been developed to provide drill or practice over materials taught in other ways, there are few programs for chemistry at the high school level that present new concepts via CAI. This research was an attempt to develop and test such materials.

Research Design and Procedure

A group of 110 students enrolled in chemistry at the University School at Florida State University were randomly assigned to treatment and control groups. The treatment group (N = 58) was given instruction on (1) structure of the atom (quantum mechanical model), and (2) chemical bonding (ionic, covalent, van der Walls, hydrogen) and molecular architecture via CAI at the CAI Center at Florida State University. The control group (N = 52) received instruction over the same topics in their regular chemistry class. During the time of the experiment, students in the treatment group reported to the CAI Center during their regularly scheduled chemistry class and had no contact with their chemistry instructor. Students in the treatment group were allowed to proceed at their own pace and had access to the CAI terminals after school hours and during study periods as well as during their regular class. Records were kept on the student responses entered, latency (time required by the student to respond to the computer), total number of correct answers, total time spent at the terminal, and any other response the student typed into the terminal after he was signed on. There were no absences in the CAI group during the study. All CAI students completed the entire program.

Students in the control group were taught the same material covered in the CAI program. This material covered four chapters in the student's text (Chemistry by Choppin and Jaffe). The control group spent 15 class periods covering the material included in the experiment.

(Reviewer's Note: There is some ambiguity in the author's report of the time involved in the experiment. At the bottom of p. 78 reference is made to "the two week period of the study" but on p. 79 it is stated that "all control group students spent three weeks covering the same material included in the CAI program." This difference may simply reflect the difference in time required by the two groups to complete the material as indicated below.)

Each student in the treatment group took a 60-item, multiple-choice test immediately upon completion of the experimental material. Students in the control group took the same test as a group in the two days following the completion of the instructional material. A 35-item test was administered to both treatment and control groups 60 days after completion of the experiment. Both tests were constructed by the author by selecting test items from the ACS-NST Cooperative Examination in High School Chemistry (1952, 1963, 1965, 1967, 1969), the Anderson-Fisk Chemistry Test (1966), and the CHEM Study Achievement Test (1963-64, 1964-65). The Kuder-Richardson 20 reliability estimates for the 60-item test were .87 and .85 for the treatment and control groups respectively. The corresponding Kuder-Richardson 20 estimates for the 35-item test were .69 and .67.

Tests of the hypotheses of no differences in mean score on the above tests were made by the t-test.

Findings

The difference in mean for the treatment group ($\overline{X}$ = 20.66; S.D. = 6.49) and the control group ($\overline{X}$ = 24.19; S.D. = 9.36) on the 60-item test administered at the close of the experiment was found to be significant at the 0.5 level. The 35-item delayed posttest showed differences in the same direction and were significant at the same level of confidence. ($\overline{X}$ Treatment Group = 10.66; S.D. = 4.64; $\overline{X}$ Control Group = 12.62; S.D. = 4.64) However, the CAI students completed the instructional material in an average time of 275 minutes compared to 750 minutes spent in regular class instruction. It was also found that students in the CAI group had a favorable attitude toward their work. (This result was previously reported and is only mentioned in this article.)

Interpretations

"The data collected in this study indicate that students learn more, as measured by posttests, with typical classroom instruction in chemistry than they do with tutorial-type computer assisted instruction." ..."Data collected in this study indicate that students can complete the same amount of material via CAI in less than one-half the time required by students in the typical classroom." ... "Strong student interest and favorable attitude toward CAI imply that this mode of instruction can be used effectively. Further, it is suggested that the difference between the mean scores on the posttests of the two groups in this study...is small. When this is compared with the positive student interest, attitude, and time-economy, the positive aspects of CAI outweigh the negative aspects." (p. 81 of article under review.)

Abstractor's Notes

This study of CAI in chemistry instruction constitutes a worthwhile addition to our understanding of CAI. It would appear that the study was carefully conducted and this reviewer sees no reason to question the internal validity of the study. The reviewer also agrees with the author that the differences in mean achievement by the treatment and control groups, though statistically significant, are small enough that one should not argue strongly for one mode of instruction in preference to another. As indicated by the standard deviations in the scores, variance within groups was considerably more than the variance between groups. Factors such as cost (not mentioned in the study), time (both developmental and instructional), and attitude (both long range and short range) must be considered along with the data on mean achievement in deciding for or against tutorial-type computer assisted instruction in chemistry.

Considerable care must be exercised in generalizing the results of this study to other instructional situations. No description of the students involved in the study is given. If the majority of the students involved in the study are bright, highly motivated students (or dull, poorly notivated students), the results may differ considerably from those that would be obtained with a different student population. It should also be noted that the study compares the effectiveness of a particular CAI program with the effectiveness of classroom instruction of a particular teacher. The lower achievement by the CAI group might be improved by an improved CAI program. The longer time required by the control group might be shortened through more effective classroom instruction. The study provides little information on the multitude of variables inherent in each instructional mode which interact to affect both achievement and instructional time. Readers who are seriously interested in CAI will certainly want to communicate with the author to obtain a copy of the CAI material for careful examination.

Perhaps the most serious question to be asked by the reader is, "How do I use the results of this research?"

Can the reader say, "Achievement will be greater under regular classroom instruction. Therefore, I should forget about CAI?" Hardly. Although this may be true, modifications in the CAI program might make it more effective or modifications (such as changing the teacher) in the classroom instruction might make it less effective, thus reversing the results found in this study. Repetition of the result found in this study with a different CAI program, a different classroom teacher, and a different group of students is problematic.

Can the reader say, "CAI instruction can be expected to take less instructional time than regular classroom instruction?" Not really. It is likely that the CAI instruction took less time because it was self-paced and the program was written rather efficiently.

But it is possible to make regular classroom instruction self-paced as well and it is possible to write CAI programs so inefficient that it takes a very long time for students to complete them. Once again, it is the unknown particulars of each mode of instruction which governs the time required for completion of the instruction.

In the view of this reviewer, comparative research of this kind is of most value to the community of science educators when it is able to tell us why the observed differences in performance and time were found, rather than that they exist. For example, based on the reviewer's own work, it is often found that differences in means such as those found in this study can be attributed to higher performance of one group on a very few items in the criterion test. (That is, item analysis of the test results often shows that the proportion of students answering each item correctly will be approximately the same for a majority of the items but may differ substantially for three or four of the test items.). It is often possible to trace this difference in performance to unintentional gaps in one or the other instructional strategy. Subsequent revision of instruction may then result in elimination of the observed differences. It is the identification of such weaknesses in instruction that will have the greatest impact on the improvement of instruction.

In similar fashion, we need to know why students were able to complete the CAI materials in less than half the time required in normal class instruction. Knowing this, it might be possible to effect the same economies using strategies that can be applied in more conventional classroom settings. If this is possible, the potential for instructional improvement is greatly enhanced.

These comments are in no way intended to demean the work reviewed here; rather, they are simply intended to suggest the kind of information that is needed in order to make evaluative studies most useful to the science education community.

EJ 049 774

Tamir, P. and F. Glassman, "A Practical Examination for BSCS Students: A Progress Report." Journal of Research in Science Teaching, Vol. 8, No. 4:307-315, 1971.
Descriptors--*Achievement, Biology, *Laboratory Procedures, *Problem Solving, Research Design, Secondary School Science, *Skills, *Tests

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

Purpose

The purpose of this investigation was fourfold in nature. It included the following problem areas:

a. a comparison of results obtained by administering a practical examination in 1969 and 1970 to Israeli twelfth grade BSCS biology students;
b. the development and use of a scoring grid to measure specific skills regardless of the fact that different problems were being administered to different students;
c. an investigation of the possible relationships that might exist among the identified skills; and
d. a determination of the practical examination's effectiveness in discriminating between BSCS and non-BSCS biology students.

Rationale

The investigation was a follow-up and extension of a previous report by the same investigators (1). It centered around the notion that pencil and paper evaluations do not measure some important student objectives that should be an integral part of a BSCS laboratory-centered approach to the teaching of biology. A further consideration was the fact that although practical examinations have been frequently stated as being a desirable component of an evaluation program for laboratory activities, they have seldom been used as evaluation tools by researchers in science education.

Research Design and Procedure

Two practical laboratory examinations for use with secondary school biology students were developed by the researchers. Each examination consisted of 15 percent plant identification with a key, 35 percent oral examination on plants and animals, and 50 percent problem solving through experimentation. One examination was administered in 1969 to 99 Israeli twelfth grade BSCS biology students as a part of their biology matriculation examination. The second examination was administered in 1970 to a similar group of 147 BSCS students. All students had received an average of four periods of biology instruction per week over a period of four years in grades nine, ten, eleven, and twelve. The data from the examinations were analyzed through the use of correlation and a comparison of mean scores.

Since the problems comprising the examinations were not identical and different students were assigned different problems, a scoring grid was devised to improve the assessment of specific student skills. Two of the skills, manipulation and self-reliance, were assessed through observation during the examination period. The remaining skills (observation, investigation, communication, and reasoning) were assessed from students' written answers. Analyses were made of intercorrelations of mean scores in the various skills.

Sixty Israeli twelfth grade biology students who had studied conventional biology for four years and who closely matched their BSCS counterparts were selected as a comparison group. The mean school grades in biology from the BSCS and non-BSCS students were 75.41 and 75.00, respectively. Two of the three teachers who taught conventional biology had received BSCS training and were teaching BSCS biology to other students. The third biology teacher had not received BSCS training but had engaged in research at the university level. Three problems from the 1969 examination and three problems from the 1970 examination were administered to the comparison group during their matriculation examination. An analysis of covariance, correlation and mean differences in performance were used to compare the scores of the 60 conventional biology students with the scores of their 142 BSCS counterparts.

Findings

The findings reported by the investigators were as follows:

a. the mean scores of the BSCS students on the 1969 and 1970 examinations were 71.76 and 73.90, respectively;
b. intercorrelations between the mean scores among the components of the practical examination revealed that plant identification with a key correlated 0.19 in 1969 and 0.17 in 1970 with problem solving through experimentation; performance in oral examination on plants and animals correlated 0.26 in 1969 and 0.35 in 1970 with plant identification with a key; performance in oral examination on plants and animals also correlated 0.44 in 1969 and 0.37 in 1970 with problem solving through experimentation; all these correlations were significant at the 0.05 level;
c. achievement on the practical examination showed a correlation coefficient of 0.23 in 1969 and 0.39 in 1970 with student scores on a pencil and paper achievement test in biology and a correlation coefficient of 0.25 in 1969 and 0.33 in 1970 with student grades in biology; these correlations were significant at the 0.01 level;
d. a scoring key was developed having manipulation (10), self-reliance (10), observation (15), investigation (20), communication (15), and reasoning (30), as the identified skills; the numbers in the parentheses represented the

weighted values of each skill in percent of the total score;

e. intercorrelations of mean scores among the various skills showed investigation correlated 0.01 with manipulation and 0.18 with self-reliance; investigation showed a significant correlation coefficient at the 0.01 level of 0.33 with observation, 0.41 with communication, and 0.56 with reasoning; reasoning also correlated significantly at the 0.01 level with manipulation (0.40), self-reliance (0.37), observation (0.45), and total examination score (0.79); and

f. the BSCS students scored significantly higher at the 0.01 and/or 0.05 levels than the conventional biology students in self-reliance, reasoning, and total examination score.

Interpretations

The following statements summarize the conclusions reported by the researchers:

a. practical examinations measure some aspects of achievement that are not measured by teacher grades or pencil and paper tests;

b. students with poor or low manipulative skills are not necessarily low in their investigative skills;

c. a practical examination has considerable impact on instruction and may be used as an agent for instructional changes in a predetermined direction;

d. BSCS biology students have an advantage over non-BSCS biology students in solving problems through the use of experiments in the laboratory; and

e. the BSCS biology course has a higher quality of concept learning associated with it than does the conventional biology course.

Abstractor's Notes

Science educators have often stated the desirability of including practical examinations into comprehensive programs of evaluation, but the statement is seldom put into practice in the science classroom. This research represents an exception. It provides an example of the development and implementation of a practical examination in biology at the secondary school level. The research further illustrates that the practical examination measured certain aspects of student achievement in biology that were not measured by pencil and paper tests or by teacher grades.

The research report is useful, but it would have been more useful to other persons wishing to develop and implement their own practical examinations if it had included a discussion of the rationale and methods used in developing the scoring grid and the procedures used in achieving interrater reliability. How were the skills selected? Why did some of the skills receive higher weighted values than others? How were the skills defined? Was it difficult to discriminate among the skills? How many practice sessions were necessary before a high interrater reliability was achieved? After a period of training, could independent raters obtain a relatively high interrater reliability? A consideration of questions such as these would have been useful because the answers are crucial for determining the objectivity of any practical examination.

The researchers came to the conclusion that the practical examinations they developed had an impact on BSCS biology instruction from one year to the next. Similar conclusions have been reached by other researchers. However, the conclusion raises some interesting areas in need of further research. To what extent can science instruction be altered through the use of practical examinations? Can the instruction be changed in predetermined directions? Are the changes more or less permanent until an examination stressing new student skills is introduced? Did the instruction in fact change, or did the students look at last year's examination and modify their own behavior independently of the instruction they received in class? Remember the examination was given as part of the matriculation examination in biology. Students are generally very sensitive to the types of questions asked on previous matriculation examinations. Systematic observation of classroom behavior might provide some evidence regarding changes in instruction. It might also be interesting to administer a similar examination the following year to non-BSCS students who had been taught by teachers who had not participated in the research nor had given the practical examination the year before as part of the matriculation examination. These questions are not to say that classroom instruction was or was not changed. They are suggesting the need for further evidence before the conclusion can be fully substantiated.

A shift in the correlation coefficients that were reported between students' total examination scores and school grades in biology illustrates another area in need of further investigation. The total examination scores of the 1969 and 1970 BSCS biology students showed significant correlations at the 0.05 level or 0.25 and 0.33, respectively, with school grades in biology. In 1970 when the practical examination was used in an attempt to discriminate between BSCS and non-BSCS biology students, a correlation coefficient of 0.59, significant at the 0.01 level, was reported between mean performance on the total examination and mean school grade in biology for the BSCS students. For the non-BSCS students the correlation between mean performance on the total examination and mean school grade in biology was 0.28, which was significant at the 0.05 level. The researchers mentioned the difference in degree of correlation when discussing the BSCS and non-BSCS students, but they did not offer an explanation for the shift from the earlier correlation coefficients. It may have been that the 1970 comparison group of BSCS teachers altered their grading criteria as a result of having become acquainted with the practical examination in 1969. The use of the scoring grid in 1969 but not in 1970 may have caused the shift in correlation. Another possibility is that yearly school grades in biology for BSCS students were quite different from mean school grades in biology for BSCS students. Hopefully, this is not the case as school grades in biology were used as the covariate in analyzing the differences in performance between the BSCS and non-BSCS students on the practical examination.

The use of matching and/or analysis of covariance can result in fairly comparable groups if relevant predictor variables, i.e. pre-measures which have an appreciable linear association with the post measures, are available. But even when such variables are available, the use of matching never assures the researcher that the differences between groups on post-measures are exclusively the result of an experimental variable. This, then, is the case in the present study. In fact, one might even question some of the predictor variables which were used. For example, matching teachers on the basis of having received BSCS training or engaging in research at the university level does not ensure that the teachers will have similar teaching methods. Before such an assumption could be acceptable, it would be necessary to systematically observe the teachers while they were in the act of teaching. Another example of a questionable variable for matching, as well as being a questionable covariate, was school grades in biology. They were linearly associated with total student performance on the practical examination, but the association was much greater for the BSCS group than the comparison non-BSCS group. In conclusion, it should be pointed out that matching and/or the use of analysis of covariance to obtain equivalent groups should be used only when it is impossible to set up a true experiment. The researcher should acknowledge his or her awareness of the fact that the use of matching and/or analysis of covariance to obtain equivalent groups does not provide a rigorous basis for assessing experimental error!

---

1. Tamir, P. and Glassman, P. "A Practical Examination for BSCS Students." Journal of Research in Science Teaching, Vol. 7, No. 2:107-112, 1970.

EJ 043 100

Tolman, Richard, "Student Performance in Lower Division Collegiate General Biology Programs in Selected Community Colleges and Four-Year Institutions in Oregon." Journal of Research in Science Teaching, Vol. 8, No. 2:105-112, 1971.
Descriptors--Academic Achievement, Biology, *College Science, *Community Colleges, *Critical Thinking, Research, *Student Characteristics, *Universities

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Robert E. Yager, University of Iowa.

Purpose

The experiment was designed to study the following questions:

(1) Is there a significant difference in student performance between four-year institutions or among community colleges in terms of sub-scores on a test of the principles of biology measuring the behavioral levels of knowledge, comprehension, application, and a combination of analysis, synthesis, and evaluation?

(2) Is there a significant difference in student performance between four-year institutions and community colleges in terms of critical thinking ability or in terms of sub-scores on a test of the principles of biology measuring the behavioral levels of knowledge, comprehension, application and a combination of analysis, synthesis, and evaluation?

Rationale

This investigation resulted from a 1967 concern in Oregon for the equivalency of educational experiences in science courses between the university system and the community college system. Notation was made of studies concerned with measuring success of community college students compared to native upper division university students. Notation was also made of a number of studies concerned with measuring the differences in attitudes, socio-economic level, and ability between students attending community colleges and those attending four-year colleges or universities.

A study conducted in New York by Kochersberger was cited as a major study concerned with reporting on the success of community college and university students in similar science courses. Kochersberger's study completed in 1965 indicated that there were no significant differences between community college and university students as measured on a common test of principles of biology. He did report that university students received more D & F grades while low-ability students performed better in the community college atmosphere.

A comparison study conducted at Oregon State University by Denney was referenced. These studies were designed to add more information concerning our knowledge of equivalency of similar courses (educational experiences) conducted in community colleges and four year universities in Oregon.

Research Design and Procedure

A total of 261 students from three Oregon Community Colleges and 465 freshmen and sophomore students from four-year institutions were selected for use in the study. All students had completed all three quarters of a general biology course and all had high school grade averages available. Campbell and Stanley's posttest only control-group model was selected as the experimental design. This design is designated:

$$R_{1,2,3,} \quad X \quad O_1O_2O_3O_4O_5$$

$$R_{4,5} \quad O_1O_2O_3O_4O_5$$

where $R_{1,2,3}$ represent the students from three community colleges and $R_{4,5}$ represent students from two universities. The experimental variable, X, was instruction in the biology course in community colleges. The students from the four-year institutions served as controls. $O_1O_2O_3O_4O_5$ represent five test scores from the populations. High school grade average and sex were held as covariates.

Four "common" objectives for the biology programs were identified. The study focused upon two of these: 1) to be able to think critically and to evaluate facts and data, and 2) to gain an understanding of the fundamental facts and principles of science. The Cornell Critical Thinking Test, Form Z, was chosen for testing of the first objective. The author chose 155 items from Testing and Evaluation in the Biological Sciences which were concerned with content in biology common to the courses at all five institutions. A critique jury consisting of an instructor from each of the five institutions assisted in limiting the test to 65 items. These items were subsequently divided into sub-areas representing knowledge, comprehension, application, and a combination of analysis, synthesis, and evaluation.

The two instruments were administered to all students (261 community college and 465 university) at their respective institutions during the same week. Both instruments were administered during the same period with a time limit of 60 minutes assigned to each instrument. All students completed each instrument in this time frame.

Findings

For the first question investigated, concerning differences between four-year institutions or among community colleges, or both, the Gauss-Markoff setup for multiple measurements (analysis of dispersion) yielded the following F-values for each test and/or sub-test:

Test	F-Value
Critical Thinking	0.37
Knowledge	6.28
Comprehension	3.36
Analysis	10.58
Combination of Analysis, Synthesis, & Evaluation	6.51

The F-values for Knowledge, Analysis, and the Combination are significant at the .01 level while the value for comprehension is significant at the .05 level.

The differences between the highest and lowest means for community colleges were compared with the standard deviations of the regression coefficients to determine the location of significant differences. The difference among the community colleges was reported to be caused by a difference between institutions on the knowledge subtest and the difference between institutions on the combined analysis, synthesis, and evaluation subtest. Similar differences between highest and lowest four-year institutions means were also compared with the standard deviations of the regression coefficients. Significant differences were found between the high and low mean scores on all four subtests of the principles of biology test (knowledge, comprehension, application, and combined analysis, synthesis, and evaluation).

For the second question investigated, concerning differences between community colleges and four-year institutions, the analysis of dispersion yielded the following F-values for each test and/or subtest:

Test	F-Value
Critical Thinking	9.35
Knowledge	29.89
Comprehension	1.11
Analysis	0.01
Combination of Analysis, Synthesis, & Evaluation	0.16

The F-values for the critical thinking test and for the knowledge subtest of the principles of biology subtests are significant at the .01 level. The F-values for the other subtests are not significant at the .05 or .01 levels. The students at four-year universities attain significantly higher scores in critical thinking and knowledge of biology than do students from community colleges.

Interpretations

The results of this study indicated that:

(1) the community college general biology students attained a level of success on three of the four subtests of the test of the principles of biology equivalent to that of the four-year institution general biology student;

(2) the community college student attained a higher level of success than the four-year institution students on the knowledge subtest of the test of the principles of biology;

(3) there was a greater difference between the four-year institutions than between the community colleges and four-year institutions on the subscores of the test of the principles of biology, and

(4) the general biology students attending the four-year institutions attained higher scores than the community college general biology students on the Cornell Critical Thinking Test, Form Z.

The author draws no inferences, suggests no implications, and does not relate the findings to the reports of other investigators. He does, however, mention some differences which were not mentioned nor discussed in connection with design. Part of the significant difference between four-year institutions appeared attributable to one additional hour per week spent in small group discussion with a biology instructor at one institution. The author also suggests that difference in critical thinking ability may be caused by the fact that students with more ability in critical thinking attend four-year colleges. The author also suggests that the reason that community college students were superior to students from four-year institutions may be that instructors in more community colleges are concerned about equivalency of courses and thereby overemphasize memorization of facts.

Abstractor's Notes

A question is raised concerning the appropriateness of the title of the paper. Since four-year colleges are held as a control group in the design and only the community colleges are being studied, a more precise title would be desirable. In a similar vein, a more precise description of relevant literature would be helpful. Also, relating back to relevant literature in the interpretation section would be most helpful.

Several comments regarding external validity seem warranted. The section of the manuscript headed "discussion" is confusing. Differences among assumptions, generalizations, explorations, and conclusions should be made. Possible explanations should certainly not be advanced as conclusions. No real "discussion" is included. The author should be more careful not to make generalizations beyond his sample. This is especially true in summarizing or stating specific conclusions resulting from the study. The statement in the "discussion" section that university students' superiority to community college students on the critical thinking test may be because university students are more capable of critical thinking is unsupported. High school grade average is covaried. Is the author suggesting that this was not an adequate covariate?

Some comments concerning internal validity are also suggested. Reliability measures of the Cornell Critical Thinking Test are low for a standardized test. Should not this fact have been noted, discussed, and considered? Reliability evidence for the biology test is desirable since it is specially constructed for this experiment. Were any data collected? The actual design used appears to be Campbell and Stanleys' static group comparisons. The posttest only control-group design, which is claimed to be used, is inappropriate for this analysis since it entails random assignment to the two "treatments" (i.e. two and four-year colleges). Motivation, selection, mortality, and interactions are all potential sources of internal invalidity. Another serious question is related to the validity of the biology examination. No construct nor criterion-related validity is reported.

Some comments related to data analysis are suggested. The criteria (i.e. cutoff level) for eliminating items in the construction of the biology test are not defined. This seems particularly important in view of the problems of the Cornell Critical Thinking Test and the results reported using the two instruments. The rationale for using the Gauss-Markoff setup for multiple measurement needs further explanation, especially since reference provides no practical information on testing hypotheses. The calculation of the degrees of freedom used seems incorrect. An explanation of why the differences between means is compared with the standard deviations of regression coefficients is also needed. The study would be greatly improved if the number of hours each course met per week were covaried. This appears to explain much of the significant differences. Further, the author at one point offers this as an explanation of difference without even identifying it as a variable initially. An explanation of why sex is covaried should be included.

With such a large sample, a measure of the degree of association (ω^2) for the F tests would be helpful in determining practical significance of results. Upon inspection, such significance does not appear to be too great. None of the subtest means vary more than one point between two-year and four-year institutions. None of the subtest means vary more than two points across all schools. With the large number of subjects used practical significance may have been masked.

Why was not a pre-test considered? A tighter design would have been possible. Assumptions concerning the initial characteristics of students enrolled at the two types of institutions would not be necessary. Obviously more extensive discussion and interpretation as well as more precise conclusions would be possible. The fact that the high school grade average for community college students were lower suggests problems. Initially the author seems to criticize other studies concerned with comparing success with general grade averages. Why did he choose high school grade-point averages (and sex!) as covariates?

Some specific questions and comments follow:

1) The notation R X 0 does not refer strictly to Campbell and Stanley criteria and is "noise" rather than "signal." Since both groups received testing, X_1 and X_2 would have been more appropriate. Since no random assignment of the students was accomplished, the author actually should have used the Static-Group Comparison (Design #3).

2) Was equivalence "proved?" (page 110 of paper)

3) Does the author use "between" and "among" in the conventional statistical manner? (see page 105)

4) The details of item analysis -- discrimination index (D) and difficulty (d) certainly need not be included in the body of the paper. A footnote reference would have provided the reader with the information without losing the flow of the paper.

5) Are Table III and IV (pages 108 and 109) needed for the general reader? They seem of little value to the study and/or to the questions raised. Perhaps greater consistency with respect to inclusion regarding descriptions, design, and analysis would be an improvement.

The limitations of the study should be more apparent for the reader. Further, the author needs to show evidence that he is aware of these limitations. Certainly the generalizations he includes lead one to question the manuscript, the conclusions, and the contributions to the field. On the other hand, many aspects of the study are clear-cut and solid. Some needed descriptions and explanations have been omitted. Other material seems to have been included for little or no reason. Making the conclusions more precise and relating them to previous reports would be a desirable improvement.