DOCUMENT RESUME

ED 111 637 SE 018 850

AUTHOR Helgeson, Stanley L., Ed.; Blosser, Patricia E., Ed.

TITLE Investigations in Science Education, Volume 1, Number 2. Expanded Abstracts and Critical Analyses of Recent Research.

INSTITUTION Ohio State Univ., Columbus. Center for Science and Mathematics Education.

PUB DATE 75

NOTE 70p.

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 second 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. The analyses are intended to provide useful comments and suggestions to serve as a device which might be useful for training in the writing of research articles. Articles included in ISE are selected primarily from such sources as professional journals and reports of government-funded projects. Abstracts included in this issue relate to topic areas such as evaluation of children's performance on the cognitive, affective and motivational domains, student perceptions of teachers, evaluation of student laboratory reports, effect of behavioral objectives, and comparison of instructional strategies. (EB)

Vol. 1, No. 2 1974-75

INVESTIGATIONS IN SCIENCE EDUCATION

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. 2, 1974-75

NOTES . . .

from the Editor

Each research report reviewed in Investigations in Science Education will include in the citation the ERIC Descriptors assigned to the original document when it was processed into the ERIC system. The reader who is interested in a particular area represented by a reviewed report can locate other documents which may be related by searching the ERIC data base with one or more of these descriptors. Identifiers, which are assigned to some (but not all) documents, are terms which are not a part of the standard ERIC vocabulary but are in common usage in some segment of the educational community. These terms may also be used to search the ERIC data base for documents related to these special topics.

Because one of the prime concerns of Investigations in Science Education is to promote better research reporting, constructive dialogue within the science education community is encouraged. Accordingly, publishable letters of response to the abstracts and analyses are invited as well as are suggestions for improving the content and format of Investigations in Science Education.

Stanley L. Helgeson
Editor

Patricia E. Blosser
Associate Editor

EJ 079 473

Allen, Leslie R., "An Evaluation of Children's Performance on Certain Cognitive, Affective, and Motivational Aspects of the Systems and Subsystems Unit of the Science Curriculum Improvement Study Elementary Science Program." Journal of Research in Science Teaching, Vol. 10, No. 2:125-134, 1973.

Descriptors--*Academic Performance, *Behavioral Objectives, *Curriculum Evaluation, Educational Research, Elementary School Science, *Intellectual Development, Science Education, *Scientific Literacy

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Kenneth D. George, University of Pennsylvania.

Purpose

The purpose of this study was to evaluate whether participation in the Science Curriculum Improvement Study (SCIS) Grade three unit, Systems and Subsystems, resulted in performance superior to those children who did not participate in the program.

Rationale

The investigator has previously evaluated the SCIS physical science units in grade one (Material Objects) and in grade two (Interaction). This present investigation is part of a larger study that will eventually evaluate all physical science units of the SCIS program.

Research Design and Procedure

The experimental (N = 87) and control (N = 89) groups were in the third grade and were the same children who participated in the first and second grade evaluations. Both groups were further subdivided by sex and socio-economic status. The investigator developed six sets of items to evaluate five objectives from the System and Subsystems unit:

1. identify variables that may affect the operation of a simple mechanical system;
2. plan an experiment in which only one variable is changed;
3. make predictions based on existing data;
4. analyze data represented on histograms;
5. recognize changes that occur during an experiment.

Each of the subjects was individually evaluated on the six sets of items, average time per subject was 25 minutes.

Findings

1. There appeared to be little difference between SCIS and non-SCIS children in respect for their preferences for school subjects.

2. SCIS children consistently outperformed non-SCIS children in handling and manipulating the materials placed in front of them by the examiner.

3. SCIS children asked more questions and made more statements about the activities than did non-SCIS children.

4. For the simplest response category (naming of objects), the performance of the two groups was almost identical. For the more "sophisticated response categories," SCIS children outperformed non-SCIS children.

5. For all situations presented, the SCIS group was able to suggest a greater number of relevant, possible variables than was the non-SCIS group.

6. SCIS children provided more correct answers than did non-SCIS children when asked questions dealing with (a) prediction from existing data, use of experimentation to obtain new data, and (b) interpretation of data from histogram.

7. SCIS children seemed to be better observers of change than non-SCIS children.

8. The responses on the first five sets of items (25 questions) were intercorrelated and factor analyzed. Three factors were extracted:

   a. motivation to explore, by handling, the object presented;

   b. recall or recognition of knowledge and the development of intellectual skills (cognitive);

   c. motivation to ask questions, or make statements, about the objects presented.

SCIS children showed a significant difference on the first two factors: motivational (exploratory) and cognitive.

Interpretations

The significance will be examined and presented at the conclusions of the six-year study.

Abstractor's Notes

1. A description of the population was included in previous articles so that the reader would have to review these before reading this article.

2. Did the control group have the same five objectives in its science program as did the experimental group?

3. If the six sets of items were specifically prepared from the Systems and Subsystems unit, didn't these items favor the SCIS group?

4. How was the validity of the items determined? What was the reliability of the items?

5. How long did the children study the System and Subsystem unit? Did the control group have as many contact hours with science as did the experimental group?

6. Was the instructor the same for both groups? If not, how many instructors were involved?

7. Who did the individual testing of the children? Was it the same person? If not, were their observations reliable? Were instructions given to the evaluators in order to assure valid observations?

8. When was the evaluation done? A simple calculation indicates over 73 hours of testing. What was being done by the children in the groups during this time? Were the children selected randomly for testing? If not, was one group tested first?

9. Most important, did the evaluator(s) know if the children were in the control group or in the experimental group? If so, how does the reader know there was no bias on the part of the evaluator?

10. What were the implications and inferences the investigator arrived at upon the completion of this study?

11. It is hoped that in publishing the results of the evaluations to be done in the fourth, fifth, and sixth grades, the investigator will attempt to include the answers to the questions posed above.

EJ 079 841

Boyd, Eunice and Kenneth D. George, "The Effect of Science Inquiry on the Abstract Categorization Behavior of Deaf Children." Journal of Research in Science Teaching, Vol. 10, No. 1:91-99, 1973.

Descriptors--*Classification, *Deaf Children, Educational Research, *Elementary School Science, Handicapped Children, *Instruction, Science Education, *Sensory Experience

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Ronald G. Good, Florida State University.

Purpose

The purpose of this research was to investigate whether or not a significant change in the level of categorization of the deaf children in the experimental group could be achieved through participation in a series of experiences in manipulating objects.

If categorization in the deaf is tied to language, its progressive development would be dependent on the development of language. No rapid change in categorization behavior could be expected, because the attainment of language is a slow and laborious achievement for the deaf. If experience is a critical factor in the development of categorization skills, a compensatory program of experiences, specifically structured toward the development of classification skills, could be expected to affect a rapid change in categorization behavior. This latter hypothesis formed the basis of the study.

Rationale

Piaget's cognitive theory posits the roots of intellectual development in the direct manipulation of the environment, not in the verbal symbol. He supports his theory that the basic cognitive structures are derived from actions with the observation that young children classify manually before they can classify linguistically. The difficulty deaf children experience in the attainment of abstract thought may be related to a dearth of experiences in the manipulation of objects, experiences familiar to hearing children because of the stimulation of verbal communication.

Education of the deaf gives almost exclusive emphasis to the teaching of language (5,7,10,12,13). Regardless of the effort exerted in this direction, the great majority of the profoundly deaf population never achieve the minimum criterion of linguistic competence, defined as the ability to comprehend and construct grammatically structured sentences (8). This combination of circumstances, the inability of a large segment of the deaf population to achieve language competence and the retardation in the development of abstract mental process in the deaf, invites the investigation of a pedagogical approach to the education of the deaf that stresses sensory experience rather than language attainment.

The goal of science education as the goal of the education of the deaf is the development of formal thought (1,2,4,6,8,11). Many science educators consider the most effective means of achieving this goal to be through inquiry. Though inquiry may be defined in many ways, it is here defined as that mode of learning whereby one discovers relationships through his own activity. This activity may involve physical interaction with the things of the environment or mental manipulation of conceptual schemes, depending upon the intellectual maturity of the inquirer.

Research Design and Procedure

Students of the Archbishop Ryan Memorial Institute for the Deaf in Philadelphia, between the ages of 10 and 13, were the subjects of the study. The lower limit of the age range was set at 10 because even under normal conditions, abstract classification skill does not usually mature before that age. All the children in this investigation were of normal intelligence, as measured by the Kuhlman-Finch Scholastic Aptitude Test, and had incurred deafness by the second year of life. The 26 children who met the criteria for selection were randomly assigned to an experimental and a control group.

The pretest-posttest control group design was used. Dual forms of the Goldstein-Sherrer Object Sorting Test were used to minimize pretest influence on posttest performance. Three cognitive styles are identified by the test: (1) concretistic, a mode of categorizing that distinguishes relationships among environmental stimuli through external features, such as spatial and temporal contiguity or perceptual similarity; (2) functional, mode of categorizing based on the external feature of use; (3) conceptual, the abstract mode of categorization, based on intrinsic attributes and essences.

Experimental treatment consisted of participation in thirty 30-minute sessions of science inquiry, structured toward the development of classification skills, and based on the physical manipulation of objects. The lessons were planned with the underlying assumption that the deaf child is different only to the extent that insufficient environmental stimulation has made him so. He knows his environment mainly through sight, smell, taste, and touch. Unaided by the categorical patterns imbedded in verbal language, he orders enviromental input by developing his own unique system of categories. The 30 sessions, developed and taught by one of the investigators, presented a structured experiential introduction to elementary classification schemes.

The sessions were held in a laboratory-classroom that was well equipped for inquiry. An effort was made to sustain an atmosphere of freedom and informality. The children were free to move around the room, use the available equipment and resources, and test materials as they wished. A free flow of communication was maintained between student and student, and between teacher and student.

Each of the inquiry sessions was 30 minutes in length. The entire program extended over a 10 week period. The control group followed the regular science curriculum of the school for an amount of time equal to that used by the 30 inquiry sessions.

Findings

One null hypothesis was tested by the research: Participation in specific science inquiry will cause no significant change in the abstract categorization behavior of deaf children, as measured by the Goldstine-Sheerer Object Sorting Test.

The two-tailed test was used and the five percent level was selected as indicative of significance. Although the study was focused on changes in abstract (conceptual) categorization, analysis of functional and concretistic categorization was made for comparative purposes.

The results of the factorial analysis lead to the following conclusions:

(1) The initial difference between groups had no significant influence in the final difference between groups.

(2) The test itself was a significant factor (educative) in the posttest change demonstrated by the experimental group.

(3) The experimental treatment effected a significant influence on the pretest-posttest change in the conceptual categorization behavior of the experimental group beyond the change effected by the educative influence of the pretest on the posttest performance.

The results of the factorial analysis of the experimental data justify the rejection of the null hypothesis that was tested. The results of the analysis demonstrate a significant change in the level of categorization used by the deaf children in the experimental group. This indicates that sensory experiences may be the critical factor in the development of categorization.

Interpretations

The analysis of pretest and posttest data indicate a significant change in compliant categorization behavior of the experimental group. The developmental gain was achieved through a program of experiences in manipulating objects. Furth's postulate that the deficient classificatory behavior of the deaf is the result of experiential restriction in early life is clearly supported by the results of this study (3).

The significant difference in conceptual categorizing behavior achieved by the experimental group is attributed to exposure to opportunities that encouraged the development of new conceptual categories. An examination of the inquiry program designed for the study reveals a strategy of continuing challenge of existing mental structures. This is a function of cognitive socialization that rarely reaches the deaf child in his daily life. A hearing child's emergence from egocentricity to socialized thought is stimulated by a gradually developed awareness that there are other ways of viewing the world than his own (9). The cognitive socialization of the deaf child is blocked by his inability to exchange viewpoints with the outside world. He can become locked into primitive coding techniques <u>because their inadequacy remains unchallenged</u>. The results of this study suggest that the diminished effectiveness of cognitive socialization in deaf children can be compensated, at least to a limited degree, by exposure to specifically structured experiential programs designed to challenge sublogical cognitive structures.

Abstractor's Notes

The authors seem to have established a definite link between manipulative opportunities for deaf children and subsequent scores on a test of classification ability. If, in fact, there is a causal relationship between certain types of manipulative activities by deaf children and their development of "abstract categorization" abilities, then the implications for curriculum decisions are clear.

The authors noted that the acquisition of language by the deaf may be related to the development of categorical (classification) ability. This possibility certainly deserves serious study and is consistent with the viewpoint that language can reflect developing cognitive structures.

One aspect of the authors' research which should be considered is the nature of the treatment. A general description of the "science inquiry" experiences was given for the experimental group, but the control group was described only as following "the regular science curriculum." Without some quantified description of the "regular" science curriculum, one is left with many questions about just what it was that was being researched. The words "science inquiry" and "regular science" have such diverse meanings that they do not really communicate with a great deal of clarity. For a study to be replicable, the classroom conditions for both experimental and control groups must be unambiguously communicated. This is a criticism that can be applied to a great many classroom studies which are reported in the various journals.

In considering reasons for the slower rate of cognitive growth for deaf children, the authors suggested that reduced sensory input causes inadequate "cognitive socialization." The deaf child is less able to "exchange viewpoints" with the environment and thus has less of a need to revise existing mental structures to better "fit" that environment. A school curriculum which is centered primarily around language experiences tends to compound an already difficult situation for the deaf child. Hearing is not needed, however, when the child interacts with manipulative materials and so the physical disadvantage of the deaf child is minimized. "Cognitive conflicts," which tend to promote the development of new mental structures, can be "built into" sets of materials for all children.

The authors' statement that deaf children can "become locked into primitive coding techniques because their inadequacy remains unchallenged" has important implications for searching out ways of challenging these children. Regular language-based means for helping children into cognitive conflict situations are even less effective than usual, when the sense of hearing is inoperative.

Results from this study tend to support the contention that cognitive growth is mainly facilitated through a child's actions rather than words. Interesting theoretical questions are raised concerning the relationship between the deaf child's cognitive growth and his/her language development. Further work is needed to explore this relationship and to determine optimum learning environments for children with restricted sensory input.

Results from this study tend to support the contention that cognitive growth is mainly facilitated through a child's actions rather than words. Interesting theoretical questions are raised concerning the relationship between the deaf child's cognitive growth and his/her language development. Further work is needed to explore this relationship and to determine optimum learning environments for children with restricted sensory input.

References

1. Advisory Committee on the Education of the Deaf, "A Report to the Secretary of Health, Education and Welfare," U.S. Department of Health, Education and Welfare, Washington, D.C., 1965.

2. Educational Policies Commission, Education and the Spirit of Science, National Education Association, Washington, D.C., 1966.

3. Furth, H. Thinking Without Language: The Psychological Implications of Deafness, The Free Press, New York, 1966.

4. Kirk, S., Educating Exceptional Children, Houghton, Mifflin Co., Boston, 1962.

5. Kopp, H., "Curriculum as Process," Volta Review, 70:376-378, 1968.

6. Leitman, A., "Science for Deaf Children," in Proceedings of the International Conference on Oral Education of the Deaf, Vol. 2, Alexander Graham Bell Association for the Deaf, Inc., Washington, D.C., 1967, pp. 1730-1757.

7. Lenneberg, E., Biological Foundations of Language, Wiley and Sons, Inc., New York, 1967.

8. National Science Teachers Association, Theory Into Action, National Science Teachers Association, Washington, D.C., 1964.

9. Piaget, J., Psychology of Intelligence, Littlefield, Adams and Company, Totowa, N. J., 1966.

10. Rosenstein, J., "Research in the Linguistic Behavior of Deaf Children," in Proceedings of the International Conference on Oral Education of the Deaf, Vol. 2, Alexander Graham Bell Association for the Deaf, Inc., Washington, D.C., 1967, pp. 1383-1392.

11. Simmons, A., "Content Subjects Through Language," Volta Review, 70:481-486, 1968.

12. Streng, A., "The Language Arts in the Curriculum for the Deaf," Volta Review, 70:487-492, 1968.

13. Woodward, H., "The Social Studies in the Schools for the Deaf," Volta Review, 70:405-409, 1968.

EJ 078 889

Bybee, Rodger W., "The Teacher I Like Best: Perceptions of Advantaged, Average and Disadvantaged Science Students." School Science and Mathematics, Vol. 73, No. 5:384-390, May, 1973.
Descriptors--Attitudes, *Educational Research, Science Education, Secondary School Science, *Student Attitudes, Student Opinion, *Teacher Behavior, *Teacher Characteristics, *Teacher Education

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Robert E. Ziegler, Elizabethtown College.

Purpose

The author of the paper, Rodger Bybee, is proposing that during the development of science education programs at colleges and universities a variety of in-puts be considered. The usual mix is composed of the present curricular materials and the experiences and perceptions of the course developer. Another in-put is being proposed from the consumer, the high school student.

Rationale

It is felt by the author of the paper that the perceptions of the high school students, who will be taught by the "products" of teacher training institutions, should be considered and reflected in the teacher training program. Three groups of students characterized as advantaged, average, and disadvantaged rated five teacher characteristics in a preferential order. The ranking of the characteristics indicates that, among the students, interpersonal relationships rank the highest within the framework, The Teacher I Like Best.

This study is related by contrast to a study by Fox and Hein that indicates college faculty perceive professional qualities highest while interpersonal relationships were the lowest.

Research Design and Procedure

Three high school student populations were given a 50 item Q-sort. The populations were characterized as advantaged, average, and disadvantaged. The advantaged group was composed of 31 students from the Frontiers of Science Institute at the University of Northern Colorado. The average group was composed of 44 students from the Laboratory School, University of Northern Colorado. The disadvantaged group of 96 students was compposed of two populations of Upward Bound students, one at the University of Northern Colorado and the second at Temple Buell College, Denver, Colorado.

The 50 item Q-sort was arranged into five major categories with 10 sub-categories under each. All of the categories and sub-categories except teaching methodology had been shown to be representative of teacher performance and were positive in reference to

TABLE 1: COMPARISON OF STUDENT POPULATION
Data are Reported in Average for Population

	Upward Bound Students	University High Students	Frontiers of Science Students
Sequential Test of Educational Progress	21.04 (15-27 Percentile)	33.0 (60-73 Percentile)	40.16 (83-96 Percentile)
Grade Point Average in Science (4 points is high)	1.8	2.8	3.9+
Grade Point Average in all School Subjects (4 points is high)	2.4	3.2	3.8
Number of School Activities	2.2	3.4	4.4
Family Income	$4,000 to $6,000	$12,000 to $14,000	$10,000 to $12,000
Number of Children in Family	5 to 6	2	3
Father's Education	Some High School	Graduated from College	Graduated from College
Mother's Education	Some High School	Some College or Trade School	Some College or Trade School

teacher behavior in a study by Cosgrove (1). On the Q-sort each item was preceded with: The Teacher I Like Best.

Administration of the Q-sort was completed in two stages: first, the students were required to sort the items into three groups, each containing approximately the same number of items. The three groups were identified as positive (17 items), neutral (16 items), and negative (17 items). Second, these groups were further separated into seven categories: Mostly strongly agree (2), strongly agree (6), least strongly agree (12), neutral (10), least strongly disagree (12), strongly disagree (6), most strongly disagree (2). Each category was allowed the number of items indicated in the parenthesis. The items were placed in envelopes indicating the category and number of responses allowed. Only acceptable student responses were analyzed, incorrect responses were rejected from the statistical analysis.

Findings

The results of the study indicate all groups of students (advantaged, average and disadvantaged) rated adequacy of relations with students in class and enthusiasm in working with students as the top items. All students ranked Teaching Methodology last. Table II presents a summary of the results.

TABLE II: RANKING OF TEACHER CHARACTERISTICS BY SELECT GROUPS OF SECONDARY SCHOOL STUDENTS
Data are Reported by Rank and Average

	Frontiers of Science Students N-31	University High Students N-44	Upward Bound Students/ Greeley N-46	Upward Bound Students/ Denver N-56
1. Knowledge and Organization of Subject Matter	3rd (3.874)	3rd (4.004)	4th (4.698)	4th (4.239)
2. Adequacy of Relations with Students in Class	2nd (3.397)	1st (3.336)	1st (3.305)	1st (3.304)
3. Adequacy of Plans and Procedures in Class	4th (4.461	4th (4.348)	3rd (4.371)	3rd (4.209)
4. Enthusiasm in Working with Students	1st (3.129)	2nd (3.373)	2nd (3.326)	2nd (3.450)
5. Teaching Method	5th (4.961)	5th (4.973)	5th (4.724)	5th (4.920)

Interpretations

Most programs for the preparation of science teachers emphasize knowledge of subject matter, ability to plan and prepare, and methods of teaching. The rankings of these students indicate that the characteristic of primary importance to them is interpersonal relationships. Development of interpersonal relationships has been assumed in programs and therefore not emphasized. Since teachers and students enter the classroom with conflicting perceptions, discord develops. The stated conflict should be confronted and attempts made to establish a higher correlation between the perceptions of the two groups.

Abstractor's Notes

The classification of students into categories such as advantaged, average, and disadvantaged is difficult any time. With a limited population of students available for this study, it is even more difficult. Therefore, it seems questionable to designate this group a "cross-section" of secondary students. It seems highly questionable to classify the University High students as average when one views the family income or educational achievement of the parents.

The procedures for the students to follow appear to be rather difficult. Of the 50 items in the Q-sort, 40 had been previously designated as positive. With this being true how can we expect students to rate the 50 items positive (17 items), neutral (16 items), and negative (17 items)? Following this original division the students are then required to arrange items into seven groups with a designated number of items in each group. The author states that incorrect responses were rejected from the statistical analysis which seems to reduce further the "cross-section" quality of the study.

"The Teacher I Like Best" certainly connotes personal qualities, therefore it isn't surprising that the categories rated most highly reflected personal qualities. If the prefix had been stated differently, such as "The Teacher From Whom I Learned The Most," the results may also have been quite different.

In the later part of the paper the author contrasts the perception of the students with the perceptions of faculty members in the study by Fox and Hein. It seems difficult to compare these studies when the prefixed statements for the two groups are different, "The Teacher I Like Best" and "Effective Teaching."

Learning frequently involves human interaction and therefore interpersonal relationships do need to be emphasized in teacher preparation programs.

References

1. Cosgrove, Don J., "Diagnostic Rating of Teacher Performance." Journal of Educational Psychology, Vol. 50, No. 5:200-204, 1959.

2. Bybee, Rodger W. and Trudy L. Dever, "The Ideal Science Teacher: A Study Comparing Perceptions of Children Pre-Service and In-Service Elementary Science Teachers." (in press)

3. Fox, A. M. and Dennis D. Hein, "Teacher Perceptions in Effective Teaching." Colorado Journal of Educational Research, Vol. 10, No. 3:45-49, Spring, 1971.

EJ 042 993

Cohen, Ronald D., "Evaluation of Student Laboratory Reports Under a Schedule of Partial Reinforcement." Journal of Research in Science Teaching, Vol. 8, No. 2:185-189, 1971.
Descriptors--Classroom Research, *Educational Psychology, *Grading, *Learning, *Reinforcement, *Secondary School Science

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Robert G. Bridgham, Michigan State University.

Purpose

This investigation focused on whether the acceptability of student laboratory reports will change when only a randomly chosen 25 percent of the reports are graded.

Rationale

The investigation was guided by a metaphor drawn from stimulus-response learning theory. In many studies (primarily of animals other than man) it has been demonstrated that performance can be maintained by providing relevant reinforcement sporadically rather than continuously. For example, Ferster and Skinner in their Schedules of Reinforcement (Appleton-Century-Crofts: New York, 1957) report a rich variety of studies in support of this principle. In the current investigation the performance to be maintained is production of an acceptable laboratory report. The reinforcement is thought to be reception of a grade on the laboratory report.

Research Design and Procedure

The group studied consisted of the 48 students in two classes of an elective ninth grade science course. The science course was available only to students with a grade of "C" or better in eighth grade science and permission of the eighth grade science teacher.

During a preliminary ten week period students were trained to write acceptable laboratory reports. Eight laboratory reports were assigned during the ten weeks, and all were graded. Any report that was not acceptable was returned to the student, who was told to revise the report and hand it in on the next day if he wished to receive credit.

In the next four weeks four more laboratory reports came due. One class (A, with 25 students) continued to have each laboratory report graded. The other class (B, with 23 students) was told that each laboratory report would be collected according to the normal schedule, but that when four reports had been collected, one would be selected at random to receive a grade. The grade assigned to that one report would be multiplied by four and would count as the grade for all four laboratory reports. A separate random selection would be made for each student. The investigator recorded the number of acceptable laboratory reports submitted by each student in both classes.

In the next four weeks four more laboratory reports came due. The grading schemes used in the two classes were swapped, with class A receiving grades by the random one-in-four scheme and class B having each laboratory report graded. Again the investigator recorded the number of acceptable laboratory reports submitted by each student.

At no time during the study did the teacher-investigator indicate that a formal inquiry was underway. Changes in grading procedure were explained as something the teacher wanted to "try out."

For the analysis, each class was divided into two groups. One group included all those whose laboratory reports were all acceptable during the preliminary ten weeks; the other group included all those who submitted one or more unacceptable reports during this period. Roughly three-quarters of the students in each class fell into the first (all acceptable) group.

The effects of the experimental treatment (one-in-four grading) were analyzed in a set of two-by-two contingency tables. Each table showed the number of acceptable vs. the number of unacceptable laboratory reports for students who had each laboratory report graded and for students who had only a random fourth of their laboratory reports graded. Four contingency tables were constructed: one for the first experimental four week period included the students whose laboratory reports in the preliminary ten weeks had all been acceptable; another for the first experimental four week period included the students who submitted one or more unacceptable laboratory reports in the preliminary ten weeks; one for the second experimental four week period included the students whose laboratory reports in the preliminary ten weeks had all been acceptable; another for the second experimental four week period included the students who had submitted one or more unacceptable laboratory reports in the preliminary ten week period.

Chi square was computed for each contingency table to determine whether the frequency of unacceptable laboratory reports was associated with the grading scheme used.

Findings

None of the four chi square values computed was large enough to approach statistical significance. The reported chi square values were each less than one.

Interpretations

The investigator concluded that "once a level of acquisition is reached, it is not necessary for a teacher to grade each response submitted by his students in order to maintain the same level of response." He notes that the results of the study are consistent with the findings that have persistently appeared in studies of partial reinforcement schedules. He also suggests that confidence in the generalizability of the study's findings may be limited by the nature of the sample and by the limited time period of the study (four weeks of the experimental condition). He comments on the difference between "acceptability" and "quality," and on the difficulty of defining specifically what was reinforcing in the complex of student behavior and teacher response that leads to the assignment of grades.

Abstractor's Notes

How far can the findings be generalized?

Among the factors that may affect our willingness to generalize from this study are the following: A) the connection of study results to well-established theory, B) idiosyncrasies in the group or situation studied, C) errors in design, execution, or reporting, and D) inadequate information about important aspects of the study.

A) The connection of study results to well-established theory - If a study's results can be interpreted in terms of a well-established principle or law, we can usually generalize from the study with more assurance. We are likely to be suspicious of generalizations to areas where the principle has proven inadequate and confident of generalization to areas where the principle has demonstrable explanatory and predictive power. The principle of partial reinforcement has not, to my knowledge, proved inadequate in any situation where performance, reinforcement, and the contingencies linking performance and reinforcement were well-defined. Thus, if the current study can be clearly connected to the principle of partial reinforcement, we should be able to generalize its findings with confidence. Unfortunately, the linkage isn't clear.

If a high ability student receives a low grade on a laboratory report is reinforcement being provided? The analysis and documentation provided by the investigator does not clearly establish that the receipt of a grade is reinforcing. Indeed, the frequency of acceptable reports in the second experimental four week period is significantly lower than that in the first experimental four week period. Apparently neither "reinforcement" regime is adequate to maintain the desired student performance. Since what defines a reinforcement is its power to maintain performance, grading may not be reinforcing. Thus, it may be inappropriate to put much weight on the resemblance of this study to studies of the effects of partial reinforcement; it is not clear that the relevant reinforcements in the situation have been identified and experimentally manipulated.

B) Idiosyncrasies in the group or situation studied - If we cannot use well-established theory to guide generalization, we are left with using our general knowledge of affairs to assess how far the findings of the study can be generalized. The fact that those students were average or above in science achievement and had elected the course becomes more crucial. If the results are trustworthy, they can probably be generalized to most other situations in which students are able and interested. It is not clear that the results can be generalized to more typical classes - those that are more varied in ability or that contain students with less interest or motivation. The limited time span of the investigation may be less critical since there was no indication that the effects of sporadic grading were shifted any more or less by the passage of time than were the effects of continual grading.

C) Errors in design, execution, or reporting - Whether or not the findings can be generalized becomes moot if the findings are untrustworthy. Are there missteps in the study or in its reporting that would make us skeptical of the results? The investigator chose the wrong statistical test. A chi square test is inappropriate when any cell in a contingency table has an expected value less than five, and each of the contingency tables in this study has such a cell. However, use of an appropriate test - Fisher's exact test - doesn't change the findings, so this error is of no consequence. More troubling is the absence of data for one student from the data tables. (Data corresponding to 47 students are presented, but, presumably, 48 students were involved in the study). Since the reporting is otherwise meticulous the discrepancy is probably a harmless oversight, though it is disconcerting. There seems to be no strong reason to doubt the trustworthiness of the findings.

D) Inadequate information about important aspects of the study - One problem in generalizing the findings of the study is that we are never told what made a laboratory report unacceptable. It is made clear that "acceptability" is different from "quality." Since quality is usually judged on the substantive content of a report, we might guess that acceptability involves a judgment about the presence of required elements in the report (e.g., are procedures, observations presented, conclusions drawn?) However, we don't know what standards for acceptability the investigator applied. Consequently we cannot be confident that we can expect similar results when we apply our own standards for acceptability.

Summary for generalizibility - If we could replicate the investigator's standards for acceptability and if our students were able and interested in their studies, we would probably get similar results from sporadic grading of student's reports. Our ability to generalize with confidence is limited by uncertainty about what is meant by "acceptability." It may also be limited by the rather special nature of the student group studied.

What is the educational relevance of the study?

At first glance the findings of the study seem to promise much. It appears that we can maintain the existing quantity of student work with less work by the teacher - a technological breakthrough. Alternatively (although this wasn't investigated) the study suggests that with the same effort that teachers now put in, we might be able to get more work from students.

Most often, though, it is not the completion of work in some formally acceptable way that is educationally relevant. A student might produce a formally acceptable laboratory report or essay or problem solution that is riddled with misinformation or misconception. Students' work <u>is</u> an occasion for reinforcement: it is, more importantly, an occasion for teacher assessment of what students know and are able to do and of the corrections of student performance and teacher planning that follow from the assessment.

While the kind of grading scheme described by the investigator may maintain the level of student work with less teacher effort, it cuts down the possibilities of educationally beneficial responses to student work. It does this by increasing the time, on average, between student completion of work and the teacher's response to the work. It also permits most student errors and misconceptions (those in the three quarters of the work that is not graded) to go unchallenged and uncorrected.

Since it is the teacher's response to student work that is most likely to bring educational benefits, and not the work by itself, the technological breakthrough suggested by this study is more apparent than real.

EJ 079 472

Gatta, Louis A., "An Analysis of the Pass-Fail Grading System as Compared to the Conventional Grading System in High School Chemistry." Journal of Research in Science Teaching, Vol. 10, No. 1:3-12, 1973.

Descriptors--*Academic Achievement, Chemistry, Educational Research, *Evaluation, *Grading, *Pass-Fail Grading, Science Education, *Secondary School Science, Student Attitudes

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Ann C. Howe, Syracuse University.

Purpose

The purpose of the study was to compare the effects of the conventional five-point grading system and the pass-fail system on (1) student achievement and (2) student attitudes in chemistry at the secondary level. The teaching practices and techniques used by teachers in the two treatments were also compared.

Rationale

The assumptions underlying the two grading systems were not explored nor were reasons given for or against either system. The assumption was implicit that readers were familiar with the two grading systems.

Research Design and Procedure

Two treatment groups were formed by random assignment of all of the 196 chemistry students in one high school to a pass-fail or conventional grading group. Each group contained four sections (classes). There were two teachers, each of whom had two sections from each group and taught one pass-fail and one conventional section in the morning and one pass-fail and one conventional group in the afternoon. The text used was Chemistry: Experiments and Principles, a revision of the original CHEM Study Course. The experiment was carried out during one semester of the 1970-71 academic year.

Student achievement was measured by two instruments, the ACS-NSTA Cooperative Examination High School Chemistry, Form 1969, Part I and Achievement Tests for Chemistry: Experiments and Principles, Series I. Four sections of the latter (Chapters 1-3, Chapters 4-6, Chapters 7-9, and Semester Exam) were administered. Data from each were analyzed separately. Student scores were divided into three achievement levels (high, middle, and low) on the basis of cumulative grade point averages. A 2 x 3 treatment-by-level analysis of variance design was used for each set of scores to test two hypotheses: (a) that there were no differences between mean scores of the two treatment groups, and (b) that there were no differences between differences in mean scores of the two groups within corresponding levels (i.e., interaction between treatments and levels).

The Scheffé test was also performed on pairs of means within each achievement level.

Student attitudes were measured by the Silance and Remmers instrument, A Scale for Measuring Attitudes Toward Any School Subject. Form B was administered as pretest and Form A as posttest. A 2 x 3 treatment-by-level analysis of variance, with pretest scores as covariate, was used to test two hypotheses: (a) that there were no differences between mean scores of the two treatment groups and (b) that there were no differences between differences in mean scores of the two groups within corresponding levels (i.e. interaction between treatments and levels).

Teaching practices and techniques were measured with the Science Classroom Activity Checklist. A 2 x 2 treatment-by-teacher analysis of variance[1] was performed to determine whether there were (a) differences in the classroom practices of the two teachers, and (b) teacher-treatment interaction.

Findings

Achievement: Analysis of the scores obtained on the various tests of achievement allowed the rejection of the null hypothesis of no treatment (main group) effect in four instances. Significant differences (.05 level) were found between mean scores of the pass-fail grading group and the conventional grading group on the ACS-NSTA Cooperative Examination, and on the Chapters 4-6 Test, the Chapters 7-9 Test, and the Semester Test of Achievement Tests for Chemistry: Experiments and Principles. In all four cases, the mean scores of the conventional grading group were significantly higher than the mean scores of the pass-fail grading group.

In addition, the null hypothesis of no interaction between treatments and levels was rejected in the case of the Semester Test of Achievement Tests for Chemistry: Experiments and Principles. The Scheffé test indicated that on this test, the high achievement level of the conventional grading group had a significantly higher mean score than the corresponding level of the pass-fail grading group. Differences at other achievement levels were not significant.

Attitude: Analysis of the data obtained from the administration Form B (pretest) and Form A (posttest) of Attitudes Toward Any School Subject allowed the rejection of the null hypothesis of no treatment (main group) effect. Students in the conventional grading group had a significantly (.05 level) higher adjusted mean score then those in the pass-fail group. No interaction was found between treatments and levels.

---

1. There is a typographical error in line 27, page 5. "A 2 x 3 analysis of variance" should be corrected to read "A 2 x 2 analysis of variance". Also, the statement of the null hypothesis in the paragraph following does not agree with the hypothesis implied in a later section.

Although no interaction was found between treatments and levels, except in one case which has been noted, the Scheffé test was performed on pairs of mean scores of all achievement tests and the attitudes test for each achievement level. In all cases, the students in the high-achievement level of the conventional grading group had significantly higher mean scores than the high-achievement level students in the pass-fail grading group. No significant differences were found on any tests for students in the middle achievement level. The results of the comparison of means for students in the low achievement level were not consistent but favored the conventional group.

Teaching Practices and Techniques: No significant differences were found between teaching practices and techniques used in the classes attended by the two groups. No interaction was found between treatments and teachers.

Interpretations

In all cases the mean scores of the conventional grading group were numerically higher than those of the pass-fail grading group. Analysis of variance showed that these differences were statistically significant at the .05 level on four achievement tests and the attitude test. In the only case in which significant treatments-by-levels interaction was found, the high-achievement level of the conventional group had a significantly higher mean score than the high-achievement level of the pass-fail group. From the results obtained from analyses of the data obtained from a variety of tests, the investigator concluded that students in the pass-fail grading group showed significantly lower achievement of course objectives and poorer attitudes than students graded on a conventional five-point grading system.

The report ends with the opinion that the results of the study indicate that pass-fail grading is not a solution to grading problems and that students like to be rewarded for high achievement and will not achieve as well if this reward is not present.

Abstractor's Notes

First, a question about methodology. The method of collecting and handling data on the Science Classroom Activity Checklist is not entirely clear. A score for each student was apparently obtained and used in the calculations, since an N of 189 is reported. It would, perhaps, have been more valuable if the focus had been on the teachers and if measures or indicators of their practices and techniques had been gathered in each class over a period of time and subjected to a different analysis. The data are needed because, although the experiment was carefully designed to control for differences between teachers, the effect of teacher bias for or against one of the grading systems could have been controlled only if the teachers had not known which system was being used in any class. Since the teachers apparently did know which system was in use, it was necessary to determine by observation whether their practices and techniques varied from class to class. As reported, it is not clear how the instrument selected for this purpose was used.

From the data presented the author seems justified in his conclusion that the conventional grading system produced higher achievement scores and more positive attitude toward the subject in the sample studied. How much weight to put on the results of the Scheffé tests is a more difficult question. Since no statistically significant interactions were found, except in the one case, purists will quarrel with the use of the Scheffé. However, the consistency of the finding that high-achievement-level students in the conventional grading group had significantly higher mean scores on all achievement tests and the attitude test, and that the scores of the middle achievement-level students were not significantly different on any of the tests is very persuasive. The probability that such consistent results are chance occurrences is low. This suggests that a study comparing the conventional system to an honors-pass-no-pass system might produce useful results. If, as may be conjectured, the high achievement-level students need a high grade as a reward but others do as well in a pass-fail system, then an honors-pass-no-pass system might motivate the high-level students and also eliminate some of the unfavorable aspects of the five-point system.

A number of other studies, most of them conducted at the college or university level, have shown that students get higher scores on achievement tests when they know that they will receive a course grade than when they are taking a course on a pass-fail basis. This repeated finding should make us reexamine the reasons for using a pass-fail system. Since it seems clear that conventional grading produces higher achievement scores, pass-fail grading can only be justified, if at all, on some other basis and the outcomes evaluated by some other means.

Another point worth considering is the motivating effect which grades apparently have, at least for high-level students. Results of this and other studies make it appear that the desire to receive a high grade motivates students to learn the material which they will need to know in order to answer correctly the questions on the achievement tests. Since we know that most of this is soon forgotten, we might profitably ask whether students could be motivated to study and learn by some other method which would produce more lasting outcomes.

EJ 049 061

Herron, J. Dudley, "The Effect of Behavioral Objectives on Student Achievement in College Chemistry." Journal of Research in Science Teaching, Vol. 8, No. 4:385-391, 1971.
Descriptors--*Achievement, *Behavioral Objectives, *Chemistry, *College Science, Instruction, *Objectives

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Vincent N. Lunetta, University of Iowa.

Purpose

Does the achievement of students in a large introductory college chemistry course improve when they are given lists of course objectives? Are there treatment (presentation of objectives) - student ability interactions?

Rationale

The author relates the investigation to four prior studies evaluating the effects of providing students with instructional objectives. He concludes that the studies have yielded little information concerning the value of course objectives to student learning. The contextual framework of the study is limited to the effects of performance objectives on student learning; the author does not relate the study to more general models or to a theoretical structure.

Research Design and Procedure

Sixteen conventionally assembled class sections of approximately 24 students each were selected for the study from the 650 students enrolled in the freshman chemistry course for science and engineering majors at Purdue University. Each of the 16 sections in the study was assigned as an experimental or as a control group. The treatment consisted of handing out six lists, containing a total of 131 performance objectives, to students in the experimental group during weeks 2, 3, 5, 7, 9 and 12 of the 16 week term. (The objectives were prepared by the investigator on the basis of course outline, assignments, past examinations, and the test; "no effort was made to make the examinations correspond exactly to the objectives lists, but ... substantial overlap did exist." Fifteen sample performance objectives are included in an appendix to the paper.)

The criterion tests were three regularly scheduled hour examinations prepared by the course lecturer. The paper does not provide information defining the weeks in which the exams were administered. Student SAT scores were used to divide the sample into three ability levels of approximately equal size. Analyses of covariance were run to determine if there were differences in mean scores for each of the ability subgroups on each of the three examinations. Analyses were also run to determine if there were differences in mean scores for the total experimental and control groups. Similar analyses were reported using "diagnostic" pretest scores as the covariate and as the basis for separating students into ability subgroups in place of the SAT data. The diagnostic test was administered at the beginning of the semester. In the data analyses the individual student was used as the experimental unit rather than class groups.

Findings

Kuder-Richardson 20 estimates of reliability for the three exams ranged from .68 - .72.

Analyses of covariance did not show significant differences in means for the first and third examinations. A significant difference favoring the experimental group at the .05 level was found for the second exam. Adjusted test means on this second exam using SAT scores as the covariate showed a difference in means for the lower third of the class favoring the experimental group at the .05 level. A difference in means favoring the experimental group at the .01 level was found for the upper third of the class on the second exam. For the middle ability group the null hypothesis was not rejected.

Using the pretest scores as covariate and as the basis for separating the class into ability subgroups, significant difference (.05) favoring the experimental group again appeared only on the second examination. In this analysis, only the top third of the class showed a significant difference in means favoring the experimental group (.05).

Interpretations

"The data for Exam 2 suggest that in a large course in which students may have difficulty in understanding exactly what is expected of them, the simple procedure of providing a list of behavioral objectives will help them learn the required material. It is not clear that the procedure is more valuable for one ability group over another ... no consistent treatment -,ability interactions were found." The investigator offers the following reasons why differences were found only in the second of the three examinations:

(1) The quality of objectives was not uniform. Objectives for Exam 3 were particularly awkward.

(2) Exam 1 is essentially a review of concepts that are normally taught in a high school chemistry course and all students in the study had completed high school chemistry. For this exam the lists of objectives added little information.

The investigator suggests that the observed differences may have been the result of more meaningful preparation by graduate instructors in charge of the experimental groups since the objectives were available to them rather than the result of more meaningful study by the students in the experimental groups.

Abstractor's Notes

This research study investigates an issue of contemporary concern and interest in education. One may question whether the author has related the study and the paper sufficiently to other relevant investigations outside the area of science education. One may also question why the study was limited to an investigation of cognitive variables. What effects did the objectives have on learner attitude and interest?

The most serious question that must be raised is whether the study clarifies the questions the author has set out to investigate. Certainly, the implications that can be drawn from the study are considerably narrower than the title of the article would imply. The significant differences favoring the experimental group that were observed in one out of three exams can hardly be generalized to a broader population in spite of the reasonably large sample size. What is the detailed evidence and rationale that enables the investigator to discriminate between the results of Exams 1, 2, and 3? The decision to use the individual as the experimental unit can also be questioned, in spite of the fact that a rationale for that decision is discussed in the paper.

The paper reports data and findings more carefully than do many contemporary papers reporting research study. In fact, the author is to be commended for his candid and explicit comments regarding the areas of weakness in the investigation. He has pointed to many of the limitations in the study that are outlined below. The following factors could well have masked the effects of the variables the investigator intended to examine.

(1) The questionable assumption that no differences existed in the teaching abilities of the graduate instructors and the methods of assignment of students to those instructors.

(2) The failure to reference exam items to specific objectives given to the students. (At least an after-the-fact review might have been conducted and selected items eliminated from the examinations on that basis. A more optimum approach would have been to use the performance objectives as criteria for creating the examinations.)

(3) The effects of the lists of objectives on the instructors of the experimental groups. (The investigator should have given the objectives to <u>all</u> of the instructors or to <u>none</u> of the instructors in both the experimental and control groups, since he was trying to assess the direct effects of objectives on student achievement.)

(4) The lack of explicit information provided students regarding the nature of the lists of objectives.

(5) The lack of control over access to the lists of objectives. (Students in the control group may well have received the lists from friends in the experimental group.)

(6) The quantity of implicit and explicit information concerning the structure and goals of the course available to the students through media other than the lists of objectives.

Although the author refers to this investigation as a "pilot study," he does not elaborate upon the implications of the study for future research. He also does not discuss the implications of the limited findings for course design or instruction.

EJ 078 924

Huff, Phyllis and Marlin Languis, "The Effects of the Use of Activities of SAPA on the Oral Communication Skills of Disadvantaged Kindergarten Children." Journal of Research in Science Teaching, Vol. 10, No. 2:165-173, 1973.

Descriptors--*Communication Skills, *Disadvantaged Youth, Educational Research, *Kindergarten Children, Language Development, *Language Skills, *Media Selection, Science Education, Socioeconomic Status

Expanded Abstract and Analysis Prepared Especially for I.S.E. by David P. Butts, University of Georgia.

Purpose

To investigate the effects of participation in the activities of Science - A Process Approach (S-APA) on the oral communication skills of disadvantaged kindergarten children.

Rationale

Science activities require the use of communication skills to allow the child to exchange ideas, identify problems, and interpret events. Thus it is logical to expect that children who lack communication skills can improve these skills by participating in science activities.

Several studies have been conducted to test this hypothesis by applying the program Science - A Process Approach. Horn and Stemmler found that with Spanish-speaking first graders, participation and concrete manipulation activities gave these children a foundation for successful learning. Ayers and Mason found that language skills related to reading readiness of kindergarten children increased with completion of Part A of Science - A Process Approach.

Research Design and Procedure

The subjects used in this study were 113 kindergarten children in an inner-city school. The subjects were enrolled in four classes, two morning and two afternoon. The experimental group consisted of a randomly chosen morning (E_1) and afternoon class (E_2). The other two classes - one morning (C_1) and one afternoon (C_2) - were the control group. The control group was made up of 48 students, 24 females and 24 males. The experimental treatment group consisted of 52 students, 23 females and 29 males.

All subjects were given a pretest -- the Test of Oral Communication Skills (TOCS). This test measures six oral transmitting (speaking) skills which include:

1) language output and expressiveness
2) vocabulary

3) general meaning and ideas
4) sentence structure
5) defining words
6) average length of sentences

The TOCS also yields two oral receiving (listening) scores including:

1) listening behavior
2) listening comprehension

The treatment consisted of 22 lessons, a through k, in the activities of Part A, Science - A Process Approach for the experimental groups. The control groups were presented with 22 lessons from Springboards to Science. The investigator taught all lessons to all groups. The lessons were presented in half-hour periods, two days per week for 12 weeks. The lessons from Springboards to Science were taught by demonstration - presentation approach and questions asked by the teacher. The teaching approach for Science - A Process Approach was guided by an inquiry-discovery technique. The children were given the materials and were free to explore and discuss their observations.

At the end of the treatment, 100 of the original 113 subjects were given the TOCS as a posttest.

Findings

Using the one-way analysis of variance, the TOCS pretest scores were analyzed and there were no significant differences among the four groups.

Relative to hypothesis one, it was found that the experimental subjects of this study performed significantly better than the control subjects in total transmitting skills and on the oral transmitting subskills of language output and expressiveness, vocabulary, general meaning and ideas, and sentence structure. Thus disadvantaged students who participate in Science - A Process Approach do give more transmitting in communication skills than similar students who participate in a different science curriculum.

Interpretation

From this study, it can be concluded that oral communication skills are clearly enhanced when SAPA curriculum is presented to disadvantaged kindergarten children.

Abstractor's Notes

This study provides meaningful fresh insights into a continuing search for better ways to assist children with learning problems. While the results are useful, the rationale of the study could be substantially strengthened if the authors had shared with the reader the reasons or the logic of expecting one science curriculum to have been useful in influencing the communication skills of five-year olds. Further, it could be questioned as to why science was selected as a subject rather than music, dramatic motion or mathematics?

As the description of the procedure is studied, the reader is told that the investigator taught the science curriculums to the experimental and control groups. The outcome of this contrast in school experiences is significantly different. The reader is left with the inference that it is due to the curriculum variable. But left to one's imagination is if this is due to specific differences in either substance or children-teacher interaction between the treatments. Or might the contrasting results be due to children's interest in the science topics--unless the topics in each program were the same. Unspecified by the authors are the variables which were contrasted in the treatment.

In reflecting on the findings of this study, one should be able to now reexamine the linkages between variables and speculate on probable causes. In this study what do the findings now tell us about how communicating skills and science instruction are related? That they appear to be related is evident. That this relationship is due to a specific curriculum, Science - A Process Approach, or the content of the science instruction, or the teacher, or the students, or the instructional materials, or the teaching strategies, or the time of day or other variables is not considered by the authors.

Finally, the authors of this study are the best qualified individuals to specify the implication of their study for instructional programs for young children. While briefly alluding to this responsibility through a "perhaps" statement, the report of their study could be strengthened by thorough discussion of what the findings of their study have to say to the teacher of young children.

EJ 076 469

Hunter, Walter E., "Individualized Approaches to Chemistry vs. Group Lecture Discussions," Journal of College Science Teaching, Vol. 2, No. 4:35-38, April, 1973.
Descriptors: *College Science, *Chemistry, Evaluation, Educational Research, *Group Discussion, Instruction, *Individualized Instruction, Science Education, *Science Activities

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Glen S. Aikenhead, University of Saskatchewan.

Purpose

The following questions were posed:

(1) How does the individualized learning approach compete with a group lecture approach with respect to general chemistry achievement?

(2) What effect does a pretest on unit objectives have on achievement?

(3) What effect do learning theory lectures have on achievement?

(4) What effect does criterion evaluation of the unit test have on achievement?

These led to the following null hypotheses:

(1) there are no significant achievement differences between students in the different treatment groups, as measured by the unit test scores.

(2) there are no significant achievement differences between students in the different treatment groups, as measured by the final test score.

(3) there are no significant interaction effects between differential treatments and the pre-existing conditions of high and low SCAT scores, as measured by student achievement on unit or final tests.

(4) there are no significant interaction effects between differential treatments and pre-existing conditions of high and low high school rank, as measured by student achievement on unit or final tests.

Rationale

No rationale was stated. No theoretical models, no previous research, and no underlying assumptions were mentioned.

Research Design & Procedure

Sample: There were five groups, each with 24 subjects randomly selected from 214 students registered in a general chemistry course at Meramec Community College.

Design: A randomized, posttest only, experimental design was used, not only to compare "individualized learning" vs. "lecture" groups but also to study three variations of the individualized learning treatment. Treatments lasted for one semester and were as follows:

Group 1 ("individualized learning") -- individualized materials (readings, audio tapes, experiments, demonstrations, films, work sheets, lab reports, seminars, and problems).

Group 2 ("pretest") -- identical to #1, plus a pretest before each of 11 units.

Group 3 ("learning theory") -- identical to #1, plus four lectures throughout the semester on how to learn chemistry.

Group 4 ("criterion unit test") -- identical to #1, plus an insistance upon at least 90 percent achievement on end of unit tests.

Group 5 ("lecture") -- similar to #1, except the individualized learning activities were replaced by two lectures covering the same concepts.

Blocking: Each group of 24 subjects were blocked in two different ways for two separate analysis: high/low SCAT and high/low high school rank.

Dependent variables: Unit tests and a final examination were used, both purporting to assess chemistry achievement (no reliability and validity data were reported).

Analysis: ANOVA was used, once for each type of blocking and for each dependent variable, followed by Tukey gap tests.

Findings

Summary of Significant Findings

Measure	Analysis	Source of Variation	Significance
UNIT TESTS	ANOVA	SCAT	.001
		H.S. rank	.025
		Treatment	.001
		Treatment x SCAT	.001
		Treatment x H.S. rank	.001
	Tukey	Groups 2 & 5	.01
		4 & 5	.01
FINAL TEST	ANOVA	SCAT	.001
		H.S. rank	.001
		Treatments	.001
	Tukey	Groups 1 & 5	.05
		2 & 5	.01
		3 & 5	.01 (sic)*
		4 & 5	.01

.05 level of probability * It should be 0.1.

In other words, the findings reported were:

(1) "True achievement differences," as measured by final and unit test scores, "exist between experimental treatment groups."

(2) "True achievement differences, as measured by final test scores, exist between groups receiving an individualized learning activities treatment and groups receiving a lecture treatment."

(3) "Significant achievement increases, as measured by final test scores, may be expected as a result of pretesting on unit objectives plus individualized learning activities."

(4) "Incremental achievement differences, as measured by final test scores, exist as a result of criterion evaluation of unit tests."

(5) "Learning theory lectures do not produce incremental achievement differences, as measured by the final test scores or unit scores."

(6) "Significant interaction exists between treatment and scholastic ability" ... and between treatment and prior academic achievement, as measured by unit test scores." (The interaction was not described.)

Interpretations

"This research confirms the hypothesis that test achievement in general college chemistry is directly related to the intimacy of student participation in the learning process ... Research results clearly indicate that students in the individualized learning pretest treatment group achieved at a higher level on the final test than the other treatment groups. The researcher concludes that practitioners can expect positive incremental differences, in final test achievement, when pretesting is used.... I conclude that the externally imposed requirements of unit test achievement are directly related to increased unit test achievement.

This research supports the utilization of an individualized approach to general college chemistry."

Abstractor's Notes

The omission of a rationale means that: (1) there is no theoretical framework from which rational hypotheses may be formulated, and (2) there is no review of similar studies in the literature which might have given a rational context for the present investigation. (A review of the literature may have mentioned studies by Grobe (2) and Vander Wal (7), both of whom essentially found null results on a greater variety of dependent variables.) Anderson's (1) plea for rationally based research could be reread to advantage by all researchers.

One basic quality of research is its ability to be replicated. The present study does not appear to be replicable because the description of the independent variables is not in operational terms. For example, what was the role of the professor? What does "lecture" really mean, in operational terms? A plethora of curriculum X vs. curriculum Y type of studies during the 1960's has not yielded reasonably unambiguous results for this very reason[Robinson, (6)]. One of the most useful methods of operationlizing independent variables is the use of a check list. Kochendorfer (3) demonstrated the feasibility of this technique. Several researchers have followed his example; for example, MacDonald (4).

The randomized, posttest only, experimental design (Campbell and Stanley's #6) is a paradigm most researchers strive to adopt. The present investigation is an exemplar in its application of such a research design. While the ANOVA should give reasonably good results, the more powerful analysis of covariance is traditionally favored as the appropriate statistic of analysis. It is also customary practice to describe the significant interactions when one discovers them.

The instrumentation is fundamental to any investigation. Unfortunately, pertinent information concerning the dependent variables is missing in the present investigation. It is essential for a researcher to report test validity and reliability data.

The researcher always walks a difficult line between overgeneralizing his conclusions and not generalizing enough from his specific set of results [Raths, (5)]. However, his logic connecting his results with his conclusions must be tight. The reader may wish to reexamine the results and the conclusions of the present study. Only conclusions #3, 5 and 6 are actually supported by the results in the summary table. The Tukey gap test revealed no significant differences among the first four experimental groups, unless a case is made supporting acceptance of 0.1 level of probability. In addition, an argument is required to logically connect "individualized learning" with "intimacy of student participation in the learning process."

A research study geared toward articulating knowledge in a particular area will have different criteria associated with its generalizations than a research study aimed at evaluating a learning experience. In the latter case, not only is statistical significance a concern, but the common sense significance should also be discussed in order to give a potential program adopter a possible basis for making a decision. That is, is it worth all the trouble of changing to an "individualized" program when one might expect the reported improvement in student achievement? In the present study, the "lecture" group is about a standard deviation below the "pretest" individualized group. Most researchers would likely agree that this observed difference likely represents a sizeable or worthwhile increment in achievement, sizeable enough to warrant the adoption of "pretest individualized learning" (assuming one could replicate the program). This issue could have been discussed by the author, thereby supporting his final interpretation that his "research supports the utilization of an individualized approach to general college chemistry."

References

1. Anderson, O. Roger. "Clarity in Reporting Research II." Journal of Research in Science Teaching, Vol. 8:201-202, September, 1969.

2. Grobe, C. H. "Comparisons in College Biology Achievement Between the Audio-Tutorial and Conventional Methods of Instruction for Non-Science Majors." Unpublished doctoral dissertation, University of North Dakota, 1971.

3. Kochendorfer, Leonard H. "Classroom Practices of High School Biology Teachers Using Different Curriculum Materials." Research and Curriculum Development in Science Education, Addison E. Lee, Editor. Austin, Texas: University of Texas Publication No. 6720, 1967.

4. MacDonald, John. "The Design, Development and Utilization of Evaluative Procedures for a Process-Approach Junior High School Science Curriculum." Unpublished doctoral dissertation, University of Alberta, Edmonton, Canada, 1974.

5. Raths, James. "The Emperor's Clothes Phenomenon in Science Education." Journal of Research in Science Teaching, Vol. 10: 201-211, September, 1973.

6. Robinson, James T. "Clarity in Reporting Research I." *Journal of Research in Science Teaching*, Vol. 6:117, June, 1969.

7. Vander Wal, J. M. "The Relationship Between Two Methods of Teaching College Biology in Achievement and Attitude." Unpublished doctoral dissertation, Western Michigan University, 1972.

EJ 073 975

Johnson, Roger, Jr., "The Process of Categorizing in High and Low Socio-Economic Status Children." Science Education, Vol. 57, No. 1:1-7, January- March, 1973.
Descriptors--*Cognitive Ability, Educational Research, *Family Background, *Learning Processes, Racial Composition, Science Education, *Socioeconomic Influences, *Student Ability, Visual Measures

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Rita W. Peterson, California State University at Hayward.

Purpose

In this study an attempt was made to examine the cognitive behavior known as categorizing among high and low socio-economic status (SES) children, and to identify strategies used by these children as they were observed in the process of categorizing.

The ability to categorize was described as an inclusive process which combines observing (recognition of attributes), comparison for differences (discrimination), and comparison for similarities (equivalence). The specific stage of categorization examined was that characterized by consistent, exhaustive, dichotomous grouping.

Rationale

Along with other mental operations, the ability to categorize has recently been accorded considerable attention due to the shift in importance from verbalized knowledge about science toward an emphasis on the cognitive processes associated with science. Gagné's model which describes the basic cognitive processes of science (observing and organizing environmental stimuli) served as the theoretical framework for this investigation. References to Inhelder and Piaget (1964) and Vygotsky (1962) allude to the developmental nature of this important cognitive behavior.

Research Design and Procedure

The population consisted of 60 kindergarten pupils randomly selected in equal numbers from four high SES classes and four low SES classes in the Oakland Unified School District of Oakland, California; chi square tables were used to determine the appropriate sample size to meet a desired explained variance level of at least 25 percent. Children were tested in small groups of four and five.

Four tests from the Lawrence Lowery Visual Resemblance Sorting Tests were administered to assess pupils' ability to categorize. Pupils were asked to categorize pictures according to shape, size and pattern using attributes singly, in paired combinations or in a triad: Test I required finding a picture that resembled an exemplar from a reservoir of pictures; Test II required findings two pictures that resembled each other within a reservoir (no exemplar); Test III required finding a picture that had no duplicate among a reservoir of scattered but matched pairs; Test IV required finding three pictures that were alike within a reservoir (no exemplar).

Individual interviews were conducted with a small group of the original sample for the purpose of studying strategies. Pupils were asked to discuss their solutions to the tests, and to replicate the tests using manipulative blocks.

Eye movements of pupils performing written tests and finger movements of pupils during interviews were also expected to provide the investigator with information about categorizing strategies used by pupils.

Analysis of Variance (Newman-Kuels Post Hoc procedures) was carried out on written test results to determine differences in performance between (1) high and low SES groups, (2) Tests I, II, III, and IV, and (3) test items varying in difficulty.

Findings

Profiles of test scores between SES groups were strikingly similar. Yet, results supported the hypothesis that high SES pupils were more able categorizers than low SES pupils. Differences between overall means for the two groups were significant ($F = 21.55$, Alpha = .05). The performance of high SES pupils was consistently superior on all tests; however, the difference between groups was not significant on Test III.

Comparisons of performance on different tests revealed that kindergarten children can find duplicate pictures in a reservoir of pictures when an exemplar is given (Test I) more easily than they can find two or three identical pictures in a reservoir when no exemplar is given (Tests II and IV) and more easily than they can exclude a picture which has no matched pair within a reservoir (Test III). Performance on Test I was significantly better (Alpha = .01) than performance on all other tests.

Pupils were found to categorize best when test items involved the attribute: shape. Percentage means were used to report test item difficulty; in descending order of difficulty, the test items were solved correctly by the following percents of pupils from the combined groups: shape - 91.7 percent, shape/size - 70.4 percent, shape/pattern - 65 percent, shape/size/pattern - 55 percent, size/pattern - 52.9 percent, size - 46.7 percent, and pattern - 35.4 percent. Differences between test items were reported to be significant ($F = 51.26$, Alpha = .05).

Descriptive data on strategies were limited, the author pointed out, due to the techniques used. In general, eye movements appeared to indicate the random choice of a picture from the reservoir and a one-by-one comparison of the picture with other pictures in the reservoir. When an appropriate match was not found, a new picture was selected and the process was repeated. This strategy was least effective when pupils were asked to find three identical pictures. After marking the first two pictures, pupils appeared to switch attributes and mark a third picture which shared some attribute with the second but not the first marked picture.

Interpretations

The superior performance of high SES pupils on categorizing tests is interpreted as evidence that "SES difference is deeper than just 'basic skills' in reading, language arts and other school tasks." The cluster of variables constituting SES, the investigator argues, "seems to powerfully influence development of mental operations required to master basic skills."

In view of his findings, the investigator recommends practice or guided experience where needed to enhance the development of this cognitive behavior. He alludes to a possible implication for elementary school science programs and pretesting of cognitive abilities. Finally, he suggests the need for a study "aimed at multiple categorizing situations and strategies over several years of development."

Abstractor's Notes

Mr. Johnson has conducted an investigation which is of interest to us as science educators because it advances our knowledge about one aspect of the cognitive process known as categorizing, as it has been observed among high and low SES kindergarten pupils.

The paper is easy to read. It is clearly anchored to a sound theoretical base (Gagné, Inhelder and Piaget, and Vygotsky). The statistical analyses are appropriate to the research design. And, the findings are presented in a reasonable and logical manner. In terms of these aspects, Mr. Johnson's research might serve as a commendable model for others.

I would like to pose some questions now which might be regarded as "minor but worth mentioning" in the interest of beginning researchers.

What is the hypothesis? The purpose of this investigation is expressed clearly and succinctly: to examine categorizing performance and strategies. However, Mr. Johnson has kept his hypothesis that high SES pupils are more able categorizers "hidden under his Introductory vest," and reveals it as an hypothesis only after he furnishes the evidence to support it in the Results. This is not a serious matter, but many readers prefer to have the author's expectations made explicit from the start. Though few would expect hypotheses to be numbered, lettered, or indented in research publications, most readers expect consistency. When initial references to hypotheses occur midway through an article, some readers pause and wonder what else they may have missed up to that point.

What is the background of the subjects? The function of this investigation was to survey the status of a particular variable, a process called categorizing, among two geographically-separated populations. In cases where the variable being assessed is not introduced by the investigator but is presumed to have developed through other influences (in this case, social and economic), readers look to the investigator to summarize all available relevant information on the comparability of the two populations sampled.

Although Mr. Johnson notes that kindergarteners were sampled "to minimize the influence of extended school experience," background on the previous experience of the subjects is limited. It would be helpful to know (a) how early in the school year the research was conducted, since instruction in categorization is part of the kindergarten curriculum in many schools, and (b) whether pupils in any of the eight classes sampled had received formal training prior to kindergarten, since many high SES pupils attend nursery schools and since some low SES pupils in the Oakland Unified School District receive enriched experiences through programs such as Head Start and funding designated by legislative actions.

What about the test and interview? A number of tests are available which may assess pupils' ability to categorize. Any normative data or information on the reliability of the Lawrence Lowery Visual Resemblance Sorting Tests would have been useful for readers interested in comparing Mr. Johnson's findings with other research.

The descriptive data on the "follow-up interviews" are extremely interesting because they hold the potential for providing important clues about similarities or differences in strategies used by high and low SES pupils. However, the description of the procedure followed during the follow-up interviews is somewhat sketchy. It would be interesting to know (a) the number of pupils interviewed, (b) the approximate duration of the interviews, and (c) perhaps even the relationship of performance on the manipulative blocks test to performance on the written test.

How are the results related to the theoretical model? In this investigation, if one assumes with Inhelder and Piaget (1964) that the ability to categorize is developmental or "begins at birth," according to Mr. Johnson, then he becomes curious about how the findings described here—specifically, on differences in difficulty between tests and test items—relate to various notions about the development of the categorizing process. The theoretical connection is possible to make, of course, but it is not too obvious in the author's Rationale or Interpretation.

Some final descriptive reference which explores how the above-mentioned results might be related to specific aspects of Piaget's developmental model, of Gagné's model on the hierarchical nature of cognitive processes in science, or of the work done by Lowery and Allen (1969) would be quite valuable. Mr. Johnson makes this kind of theoretical connection very effectively with a brief reference to Deutsch and Brown (1963) and SES.

EJ 079 840

Linn, Marcia C. and Rita W. Peterson, "The Effect of Direct Experience with Objects on Middle Class, Culturally Diverse and Visually Impaired Young Children." Journal of Research in Science Teaching, Vol. 10, No. 1:83-90, 1973.
Descriptors--Ability, *Classification, *Educational Research, *Elementary School Science, *Instruction, Science Course Improvement Project, *Science Education, Student Ability

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Delmar Janke, Texas A & M University.

Purpose

The purpose of the study was to investigate Piaget's theory of equilibration by comparing the effect of direct experience with objects on the logical reasoning ability of middle class (MC), culturally diverse (CD), and visually impaired (VI) children.

In one experiment a Piagetian task was used to compare the classificatory ability of MC, CD, and VI groups of children studying SCIS Material Objects with those who had not studied the unit.

Another experiment measured MC and VI students' ability to name properties and materials and to sort objects.

Rationale

The investigators identified several earlier studies which compared the logical reasoning ability of children from different environments.

"Bovet (1) and Bruner (2) studied children without school experience and found they learn conservation of substance and liquid more slowly than do children who attend school. Culturally disadvantaged children progress at a slower rate on classification than do middle class children (3), (4). Hayes (5) and Newland (6) have shown that visually impaired children perform more poorly than sighted children on tests of intelligence. Piaget (7) reports that blind children, because they lack visual experience with objects, develop logical abilities more slowly than sighted children. Classification tasks have been used in many of Piaget's (8), (9) studies of logical reasoning. Several recent elementary school curricula include units to give children experience in classification (10), (11), (12)."

Research Design and Procedure

A. First Experiment

The first experiment involved 197 first grade students -- four CD classes, five MC classes and one VI class. The experimental group consisted of seven classes (four MC, two CD, and one VI) where

SCIS Material Objects had been taught. The control group consisted of one MC and two CD classes. The VI class served as its own control due to the small number of subjects (N = 15).

The investigator discussed with each class what it meant to float and what it meant to sink. Then each student was given ten objects on a tray and asked to sort the objects into two piles -- those which he thought would float and those which he thought would sink. The student was then asked to test his predictions.

Following this the investigator individually interviewed each student using six questions adapted from Inhelder and Piaget. Finally, each student was assigned to one of three stages of development similar to the stages described in Inhelder and Piaget (3).

B. Second Experiment

The second experiment involved 30 first grade students -- 15 MC students and the same 15 VI students who were tested in experiment one. In this experiment all students served as their own controls.

The investigators devised a test of ability to (1) name the classes that an object belongs in (i.e., to name the materials and properties of objects), and (2) sort objects. Each child was given eight objects and asked two questions of each object. He was first asked, "Here is the first (etc.) object. Can you describe this object?" After he had answered that question he was asked, "What else can you tell me about it?" The objects were then placed in a tray and the students did a sorting exercise.

The investigators listed and categorized all of the descriptors in the first part of the experiment into three categories: properties, preproperties, and others. For the sorting task, the investigators categorized the students' reasons for forming the groups according to the stages described by Inhelder and Piaget (9).

Findings

A. First Experiment

The investigation revealed that experientally disadvantaged first graders (both CD and VI) who had studied Material Objects scored significantly higher on the floating and sinking exercise than did their controls. The MC students scored very high on the exercise whether or not they had studied Material Objects.

Scores for the VI and CD students who had studied Material Objects did not differ significantly from each other on the pre or post - tests. However, on both the pre and post - tests they differed from the MC students.

B. Second Experiment

In the "naming the classes" portion of the experiment both the MC and the VI students used significantly more properties on the posttest than on the pretest. The VI students also used significantly more preproperties and total descriptors on the posttest than on the pretest. The VI students used significantly fewer properties and preproperties than the MC students on the pretest while on the posttest there were no significant differences.

Results from the sorting exercise revealed that the study of Material Objects improved the performance of both the MC and VI students. There were no significant differences between the two groups on either the pretest or the posttest.

Interpretations

The investigators drew the following interpretations from the study:

"This study shows that direct experience with objects can lead to increases in classification ability. Also these experiences are more relevant for VI and CD children than for MC children. On all of the measures used in this study, VI and CD children had similar scores. One explanation of this similarity in scores is that both groups of children come from experientially limited environments. Thus, these results support Piaget's theory that logical thinking, which includes classificatory ability, is fostered by interaction between the individual and his environment. Furthermore, direct experience with objects was effective in increasing the classification scores of children from limited environments more often than it was effective in increasing the classification scores of children from middle class environments. This indicates that for six- and seven-year-old children who initially score below the MC level, a relatively small amount of directed experience can bring them closer to the MC and in some cases up to the MC level."

Abstractor's Notes

This investigation is one of a large number which needs to be conducted to determine the viability of theories which attempt to explain how children learn — in this case, attention is given to parts of Piaget's theory.

The study does support parts of Piaget's theory. However, as the authors indicate, neither intelligence nor age was strictly controlled in the investigation and therefore the differences between groups are open to question. In any case, the differences within groups do support Piaget's theory.

An interesting question arises due to the method of selecting classes for the study. The criteria for selecting classes were (1) the teacher's reputation of excellence as described by her peers and supervisors and (2) classroom observations that indicated children were given freedom to explore within the curriculum structure. That question is: "Was the lack of difference within the experimental MC groups on experiment one completely or partially attributable to the nature of learning experiences already present in their classrooms?" And, if the answer to this question is "yes"; "What were the nature of those learning experiences?"

Because there were no control classes for experiment two, it is not clear that direct experience with objects was responsible for the increase in score on the posttest. It is possible that other types of experiences or that maturation alone could account for the increase in score.

Evidence from this investigation indicates that MC children have skill in classification before having direct experience with objects in the Material Objects unit. The evidence also indicates that 15-20 hours of direct experience significantly improve classification skill of CD and VI students although they remain less skillful than MC students. There is some question, as the investigators indicate, whether the CD and VI students have fewer opportunities to interact with the environment than MC students have or whether they have qualitatively different interactions with the environment than do MC students. Thus, it might be that the kinds of direct experiences provided in Material Objects was more important than just providing direct experiences in raising the posttest scores for the CD and VI students in Experiment One.

In experiments such as those conducted, one is often tempted to wonder whether or not the pretest is a significant contributor to any increase in score on the posttest. This is one area of investigation which might be considered more often in many research projects.

References

1. Bovet, M. C., "Etudes interculturelles du developpement intellectuel et processue d'apprentissage." Revue Suisse de Psychologie Pure et Appliquee, 27:189-199, 1968.

2. Bruner, J. S., "Studies in Cognitive Growth. John Wiley & Sons, Inc., New York, 1966.

3. Almy, M., et al. Logical Thinking in Second Grade. Teachers College Press, Columbia University, New York, 1970.

4. Wei, T., C. Lavatelli, and R. S. Jones, "Piaget's Concept of Classification: A Comparative Study of Socially Disadvantaged and Middle Class Young Children." Child Development, 42:919-927, 1971.

5. Hayes, S. P., "The New Revision of the Binet Intelligence Test for the Blind." The Teachers' Forum, 2:2-4, 1929.

6. Newland, T. E., "The Blind Learning Aptitude Test." U. S. Office of Education, Washington, D.C., Final Report OEG-3-6-061928-1558, 1969.

7. Piaget, J., Science of Education and the Psychology of the Child. Orion Press, New York, 1970, p. 30.

8. Inhelder, B. and J. Piaget, The Growth of Logical Thinking from Childhood to Adolescence. Basic Books, Inc., New York, 1958.

9. Inhelder, B. and J. Piaget, The Early Growth of Logic in the Child, Harper & Row, New York, 1964.

10. Science Curriculum Improvement Study, Material Objects. Rand McNally & Company, Chicago, 1970.

11. AAAS Commission on Science Education, Science - A Process Approach, Part A (Classifying), Xerox Education Division, New York, 1967.

12. Educational Testing Service, Let's Look at First Graders. Educational Testing Service for the Board of Education of New York, New York, 1965.

EJ 079 838

Nelson, Miles A., "Discussion Strategies and Learning Science Principles." Journal of Research in Science Teaching, Vol. 10: No. 1:25-38, 1973.
Descriptors--*Discussion Experience, *Educational Research, *Elementary School Science, Instruction, Questioning Techniques, *Science Education, *Teaching Techniques

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Dorothy L. Gabel, Indiana University.

Purpose

The purpose of this investigation was to determine the effect of two types of post laboratory discussion strategies on sixth grade students' learning of selected science principles. Both strategies involved children reporting their observations to the class. One strategy elicited inferences from students by pointing out conflicts in their observations (Probing Technique) whereas teachers using the other strategy explained the observations to the class (Nonprobing Techniques).

Rationale

This study is based on the Piagetian theory that, in order to acquire knowledge, there must be active interaction between a learner and his environment. This involves direct experience (the laboratory approach) and a mentally active learner (as can be obtained through peer interaction utilizing cognitive conflicts of Smedlund and Inquiry Training procedures of Suchman). Hence, according to both Piaget and Bruner, both laboratory experiences and discussions are important in learning.

It has been Nelson's observation that many teachers who are using the new science curricula fail to utilize discussion techniques after the laboratory activity has been completed to bring about achievement of the program's objectives. The post laboratory discussion frequently involves reporting of the observations by the students and conclusions by the teachers during which time the children are passive. If after students report their observations, the teacher elicits inferences from the children by pointing out conflicts in their observations, the children should be more mentally active and more learning should occur.

Research Design and Procedures

The sample consisted of eight classes of sixth grade students in two Philadelphia area schools. One of these was classified as urban (M.A. = 82) and the other as surburban (M.A. = 111). Two units from the science program (which consisted of topics suitable for sixth grade students and which were designed to provide practice for students with the process skills) were selected for the teaching of the science principles. Classes were taught for three forty-five minute periods per week and the teaching of the selected principles did not occur until six weeks after the experiment began. In order to measure student learning of the selected principles, two multiple choice-type tests based on the objectives of the units were constructed.

The units were taught by two experienced teachers. Each taught four classes (two in each school) using each strategy. Strategies were randomly assigned to classes. This resulted in a 2 x 2 x 2 factorial design with teacher, school, and strategy as the factors. Data were analyzed using an analysis of variance with fixed factors. Classrooms were considered the experimental unit. This resulted in using class mean gain scores for an analysis of variance with one observation per cell. Because of this, the analysis utilized the triple mean square as an estimate of the variance of group means.

The two strategies which were used in the classes were broken down into three stages. The probing strategy consisted of a data gathering phase, a data processing phase, and a verification phase. The nonprobing strategy consisted of an identical data gathering phase but this was followed by two lecture phases. In order to assure that the treatments were applied as designated, the Classroom Observational Record (COR) was employed to determine the extent to which each post laboratory discussion strategy was used. Data were recorded for four observations (visits) of each classroom and analyzed by a 2^3 analysis of variance.

Findings

Analysis of the COR data revealed significant differences at the 0.01 level in the direction implied by the discussion strategies (tables were not reported for this analysis). In addition, the data were examined on the microscopic level for the frequency of occurrence of behaviors associated with each type of model and were then analyzed by chi square. This analysis revealed that the probing strategy used significantly more data processing interaction, whereas the nonprobing strategy had significantly greater lecture and data gathering interaction. In the case of verification, however, the probing technique had only one interaction as compared with zero interactions for the nonprobing technique, therefore, the strategies did not differ in this regard.

Class mean gain scores showed no significant correlation with class mean pretest scores, lending support to the inference that regression to the mean may not have affected the outcomes. An ANOVA performed on the class mean gain scores revealed a significant difference between each strategy which favored the nonprobing strategy. The factor "school" showed the suburban school gains to be significantly greater than the urban school gains.

Interpretation

This study suggests that the type of the post laboratory discussion affects the learning of science principles. In using a probing strategy not as much content can be covered in the same amount of time as in classes taught by a nonprobing strategy. (In the probing classrooms students heard at least four applications of the principle whereas in the nonprobing classroom only one or two applications were discussed.) This may have been the reason why the nonprobing classes achieved significantly higher on a science principles test.

The study also indicated that children in the inner city school achieved significantly lower than did children in the suburban school. This may be due to the nature of the test. Even though the test was read to the students so that the urban children would not be penalized due to lower reading level, the students may not have understood the meaning of the words. A correlation at the 0.05 level between mental ability scores and the test gain scores indicated that gain may have been due to mental ability. In addition, suburban children may have been more motivated to do well on tests.

This report suggests that the newer elementary science curricula which stress the products and processes of science cannot both be learned using the same discussion techniques. If the primary emphasis is to teach principles, a nonprobing technique is to be preferred.

Abstractor's Notes

Several features of the study were excellent. It appears to have been very well designed.

1. It was conducted over a long period of time (11 weeks). This allowed classes to become adjusted to the discussion strategy.

2. The teaching was done by the same two teachers in two different schools, that is, urban and suburban. This eliminates teacher effect and allows one to look at the effect among children of different socio-economic backgrounds.

3. Care was taken to assure that the treatment was actually occurring in the classroom through the use of the COR and analysis of the observations.

Several areas were excluded from the reporting of this study which would be of interest to the reader.

1. The absence or presence of homogeneity of variance is not noted. There might have been a ceiling effect on the test for the suburban children.

2. The actual pre and posttest scores were not reported although the class gain scores were. It might have been interesting to examine the pre and post scores of the urban and suburban children in order to examine relative gains.

3. The handling of the missing data was not included. Absenteeism is frequently higher in urban settings. If this was the case and these data were omitted from class means, there may be even a greater difference between urban and suburban scores than was reported.

The statistical design of the study was ANOVA with strategy, teacher, and school as fixed factors. Because school and teacher are considered as fixed factors in the analysis, care must be taken in generalizing to other teachers and schools.

In addition, the study could have been strengthened by increasing the number of observations per cell to more than one. This would have resulted in not having to use the triple interaction mean square for the analysis and one would not have to be concerned if the assumptions for using this method were met.

Although the author states that there may be many factors which led to lower mean gains for urban children than for suburban children, replication of this study in which mental ability is used as a covariate may answer the question. There is an average 30 point difference in mental ability between the two groups. Perhaps a study of these two strategies in a suburban school between high and low mental ability children should be undertaken.

One of the reasons given for discovery learning is that children retain information and principles more than when children are taught in a didactic manner. This study does not set out to answer the question of how children would retain the principles according to which type of post laboratory discussion they had. The answer to this question is of more educational significance than are the immediate outcomes. Additional research in this area should investigate long term effects.

EJ 049 188

Ring, Donald G. and Joseph D. Novak, "The Effects of Cognitive Structure Variables on Achievement in College Chemistry." Journal of Research in Science Teaching, Vol. 8, No. 4:325-333, 1971.

Descriptors--*Chemistry, College Science, *Cognitive Processes, *Educational Psychology, Educational Research, *Learning Theories

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Heidi Kass, University of Alberta

Purpose

The study determines the relative influence on facts and subsuming concepts possessed by beginning college chemistry students of subsequent achievement in a chemistry course.

Rationale

Ausubel's model of meaningful verbal learning provides the theoretical framework for the study which focuses on the assimilation of new information into existing cognitive structure. The premise tested is that existing cognitive structure facilitates the learning of new material to the extent that it provides the relevant subsumers for the new material. The study was designed so as to meet the requirements of Ausubel's long-term transfer paradigm.

Research Design and Procedure

A sample of 770 students enrolled in their first course in college chemistry at Cornell University was administered a Cognitive Structure Exam (CS Exam) prior to instruction in order to determine the levels of prior factual knowledge and subsuming concepts of each subject. The CS Exam consists of twenty-six items selected from a "Test of Specific Course Objectives" prepared by Educational Testing Services and revised by the authors. Items measuring the presence of information considered to be somewhat arbitrarily related to concepts in chemistry were categorized as fact items. Items requiring comprehension and/or application of chemical concepts were categorized as subsumer items. Split-half reliabilities with application of the Spearman-Brown formula yielded reliability coefficients of 0.51 for the fact items, 0.67 for the subsumer items, and 0.90 for the total test.

The dependent variable of chemistry achievement was a composite score consisting of the sum of the individual test grades obtained by the subjects on tests devised by the General Chemistry Faculty at Cornell and administered at periodic intervals as the course progressed. Type of high school chemistry preparation, SAT scores, and selected attitude measures were obtained from student records and from a questionnaire.

Analysis of covariance was used to compare high, medium, and low fact groups on chemistry achievement with SAT Verbal and SAT Mathematics scores as covariates. The procedure was repeated for high, medium, and low subsumer groups. The interaction of fact and subsumer pretest performance on achievement was examined by analysis of variance which compared the achievement of a group of subjects who had high fact-high subsumer scores or low fact-low subsumer scores with a group who had either high fact-low subsumer or low fact-high subsumer pretest scores.

Findings

There was a significant relationship between pretest factual knowledge and course achievement. There was a significant relationship between the extent of prior existence of subsumers or organizing concepts in the cognitive structure and course achievement. A significant interaction between fact and subsumer performance with respect to subsequent course achievement was found. The presence of a high level of facts in conjunction with relatively few organizing concepts had little facilitating effect on subsequent course achievement. Highest achievement occurred when both a high number of facts and a high number of subsuming concepts were present.

Interpretations

The findings support Ausubel's theory of learning in that the facilitating effect of facts on subsequent learning of new course material was dependent upon the presence of organizing concepts or subsumers. Factual knowledge in itself does little to increase the efficiency of subsequent learning as in the absence of relevant subsumers new material is processed as rote information. A highly differentiated cognitive structure in which both facts and concepts are present is optimal for subsequent learning.

Abstractor's Notes

It is difficult to refute on common sense grounds alone that what the learner already knows about the subject is an important factor in determining his performance in subsequent learning of the subject. In other words, earlier success is a good predictor of later success in a given domain. Since the students in the sample had already experienced some degree of success in their high school chemistry and possessed a knowledge of chemistry at least sufficient to qualify them to enroll in a college chemistry course, one may suggest that the role of concepts vs. facts in subsequent course achievement may depend, among other things, on how new the "new" material is in relation to the basic principles of the subject. Introductory undergraduate chemistry courses often deal with topics introduced in high school. Thus the fact-concept selection and relative emphasis of the course and its achievement tests and similarity with the fact-concept emphasis in previous instruction may influence the relationships observed. For example, students with a good memory may and often do perform well on chemistry tests with a high recall component.

This raises the question of the point in learning a subject at which systematic development of concepts should begin. Studies similar to the above conducted at both the initial stages and more advanced stages of learning a subject could be revealing. What is the nature of a relevant subsumer for students introduced to a subject for the first time? Is it a verbal passage at a higher level of abstraction than the material to come, or is it a concept from the subject? If so, what sort of a concept, and what should its relation be to the facts, if any, at hand? Could an analogy function as a subsumer? Ring and Novak use "subsumer," "subsuming concept," "organizing concept" and "concept" interchangeably and, from the example given, seem to mean the concepts and principles of the subject matter. The work of Piaget and others suggests that certain logical operational schema may also serve as cognitive structure variables or subsumers independent of the specific subject-matter.

The fact-concept relationship itself warrants further examination not only from the inductive vs. deductive aspects of their presentation to learners but also from the standpoint of establishing the learning consequences in terms of variables such as efficiency and retention of various combinations and proportions of each. Depending on the topic, knowledge of certain "facts" could make learning more efficient. In introductory chemistry most students learn quite quickly the "fact" that the atomic weight of oxygen is 16.00 because it is more efficient, in terms of its frequency of use, to know this than to have to look it up every time its use is required. (The word "fact" is in inverted commas because it is not always easy in the sciences to establish what is a fact and in what sense it is a fact. As pointed out by Schwab, a statement about the properties of electrons or neutrinos is not the same kind of statement as one specifying the color of the car in front of the house.)

The designation of a fact item as one ascertaining the presence of information arbitrarily related to concepts in cognitive structure leaves the role of relevant facts, i.e. ones which are related to given concepts, in need of further clarification. Being able to write the correct electron configuration for magnesium (categorized as a fact item by Ring and Novak) may reflect the result of arbitrary memorization of rules for orbital filling. It may also, however, be the consequence of possession of concepts or subsumers relating energy sublevels, shapes and orientations of orbitals, and electron spin to atomic number. In what sense, then, are scientific concepts subsumers in cognitive structure? For example, are relational concepts more efficient as subsumers for new learning than constructs such as energy level or atomic orbital? To what extent is cognitive structure isomorphic with the structure of the subject?

The implications for curriculum design and kind of instruction of a more complete understanding of cognitive structure variables such as facts and concepts may do something to resolve the cycles of complaint that high school chemistry programs are either too theoretical or not theoretical enough. Does the student who can answer complex questions about solubility products but is not too sure what silver chloride looks like necessarily have the deeper understanding of chemistry?

EJ 079 138

Seymour, Lowell A. and Frank X. Sutman, "Critical Thinking Ability, Open-Mindedness, and Knowledge of the Processes of Science of Chemistry and Non-Chemistry Students." Journal of Research in Science Teaching, Vol. 10, No. 2:159-163, 1973.
Descriptors--*Chemistry, *Critical Thinking, *Curriculum Evaluation, *Educational Objectives, Educational Research, Instructional Materials, Science Education, Secondary School Science, *Student Attitudes

Expanded Abstract and Analysis Prepared Especially for I.S.E. by William Torop, West Chester State College.

Purpose

Two null hypotheses were tested:

1) "Experimental group one's posttest scores will not be significantly different from their pretest scores on the tests administered to measure critical thinking ability, open-mindedness, and knowledge of the processes of science."

2) "The posttest scores for experimental group one will not be significantly greater than the posttest scores of control group two on the tests administered to measure critical thinking ability, open-mindedness, and knowledge of the processes of science."

Rationale

In 1968 the Chemistry Curriculum Committee of the Philadelphia Public School District developed a guide for use by teachers in the comprehensive high schools. Three objectives for the guide were to develop critical thinking, open-mindedness, and knowledge of the processes of science. This particular study is related to previous investigations by the documentation of writers who considered these three intellectual skills to be worthwhile objectives of science instruction.

Research Design and Procedure

The experimental design appears to be the Solomon four-group design. Group one was the "experimental group comprised of eleventh grade chemistry students that were pre- and posttested." Group two was the "control group comprised of eleventh grade non-chemistry students that were pre- and posttested." Non-chemistry students were enrolled in all the same types of courses as the experimental group, except for chemistry. The population consisted of all eleventh grade chemistry students in 16 comprehensive high schools. The study itself utilized two representative comprehensive public high schools in Philadelphia during the 1969-70 school year. The eleventh grade students in these two schools were similar to the students in the 16 schools in racial distribution and academic ability as measured by the Cooperative School and College Ability Tests - Series II.

The variables were defined by the instruments used in this study. Critical thinking ability was measured by the Watson-Glaser Critical Thinking Appraisal - Form YM; open-mindedness was measured by Rokeach Dogmatism Scale - Form E; and knowledge of the process of science was measured by the Wisconsin Inventory of Science Processes. The results of pretesting indicated that there were no significant differences between the experimental and control groups in initial abilities in the three variables.

Findings

The results of testing null hypotheses one and two concerning critical thinking ability led to their rejection at the 0.05 level of significance. Chemistry students did develop their ability to think critically and were significantly superior to non-chemistry students when posttested.

Concerning open-mindedness, null hypotheses one was not rejected and null hypotheses two was rejected at the 0.05 level. This indicates that chemistry students are more open-minded than non-chemistry students.

With knowledge of the processes of science, null hypothesis one was rejected and null hypothesis two was not rejected at the 0.05 level. Chemistry students did not increase their knowledge of the processes of science.

Interpretations

Critical thinking ability: The percentile equivalents obtained by above-average chemistry students in this study were below the national values for above-average intelligence chemistry students, more than half were below the 44th percentile. "Even though instruction in chemistry appeared to improve the development of critical thinking, there might well be additional experiences in school, especially in chemistry classes ... to aid in this development."

Open-mindedness: Although chemistry students are more open-minded than non-chemistry students, the evidence is inconclusive as to whether open-mindedness is being developed through instruction in chemistry, following the guide.

Knowledge of the processes of sciences: The results of this study lead to the conclusion that the guide and its implementation totally fail in this area. A re-examination of the guide, teaching methods and philosophy is suggested.

Abstractor's Notes

The first problem was to establish the actual experimental design. The reference cited in the article, Sax (2), is actually based on the classic Campbell and Stanley chapter in Gage's Handbook (1). Even though a secondary source is cited, the terminology, itself, does not indicate the true experimental design used. The design was inferred from the fact that the null hypotheses mention only experimental group one and control group two, as well as the section on definition of terms. However, in describing the sample selection, the authors refer to experimental groups one and three. This fits their statement concerning their determination and limiting of pretest sensitization. That is, one experimental group was given the pretests while the other was not and one control group was given the pretests while one was not, for a total of four groups: two experimental and two control. Although the authors state "only the results of two groups are provided because of space limitations," reference to another group, two paragraphs later, leads to ambiguity. The Solomon four group design was finally confirmed by a telephone call to one of the authors (3).

This telephone call also provided the information that this study was only a preliminary analysis for research that has continued through the Israeli Science Teaching Center, Hebrew University, Jerusalem. However, there is no mention of this in the article itself.

While the two schools used in the study itself were similar to the city average of 16 schools in racial distribution and academic ability, there is a great amount of individual variation between individual Philadelphia schools in these two categories. Subsequent extrapolation of results to the entire eleventh grade population of the 16 schools may not be warranted.

In a similar manner there appears to be the assumption that the Chemistry Curriculum Guide was followed as written, at least in the two schools used in the study. There is no evidence to indicate similar experiences for the chemistry students involved in the study. The authors do suggest, for future considerations, examining the means by which the Guide is implemented in the classroom as well as methods of teaching and the philosophy of the chemistry program itself.

The control group was described as eleventh grade non-chemistry students. However, there is no indication given as to how much science they had previous to the eleventh grade or whether they were taking any other science courses that year. Perhaps the differences found result from a selection factor of these eleventh grade chemistry students who usually have biology in the tenth grade and physics in the twelfth grade, and the lack of differences found result from a confounding of the chemistry variable by other science courses during the eleventh grade.

It was stated in the article that "intelligence was measured by the Philadelphia Mental Ability Test," but how the results were used was not reported, at least in the article itself. Pretesting indicated that the groups were comparable in critical thinking ability, open-mindedness and knowledge of the processes of science.

Null hypothesis one was analyzed and reported as an F ratio, and null hypothesis two with a t-test. From the article itself, which states the first null hypothesis as a comparison of posttest and pretest scores for experimental group one, it would appear that a simple, repeated measures t-test would also be warranted. However, if the effect of pretest, intelligence, or some other factor is being controlled for in the statistics to warrant use of the F ratio, it is not evident in the article as published. Likewise, the synopsis reports that critical thinking ability and open-mindedness appears to have been developed but only when pretesting was part of the teaching process (emphsis mine). It is unfortunate that the reported "space limitations" did not allow for amplication of this point.

Finally, the study lasted for just one semester. Perhaps a longer time period, as well as other approaches of meeting the objectives of high school chemistry instruction, would produce more conclusive results.

References

1. Campbell, Donald T. and Julian C. Stanley. "Experimental and Quasi-Experimental Designs for Research on Teaching," Chapter 5 in Handbook of Research on Teaching, N. L. Gage, Ed. Chicago: Rand-McNally, 1963, pp. 171-246.

2. Sax, G. Empirical Foundations of Educational Research. Englewood Cliffs, New Jersey: Prentice-Hall, 1968.

3. Sutman, Frank X., Personal communication.

EJ 037 868

Wasik, John L., "A Comparison of Cognitive Performance of PSSC and Non-PSSC Physics Students." Journal of Research in Science Teaching, Vol. 8, No. 1:85-90, 1971.
Descriptors--*Cognitive Development, *Cognitive Processes, *Knowledge Level, *Physics, *Secondary School Science, Evaluation, [College Entrance Examination Board Physics Achievement Tests, Physical Science Study Committee, Scholastic Aptitude Test (SAT)]

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

Purpose

To determine what relationships exist between the level of cognitive performance of students and their high school physics curriculum background (PSSC or non-PSSC).

Rationale

Many attempts to evaluate the "new" science curricula in the past have been inadequate due to the lack of criterion measures which are not biased toward the content emphasized in a particular curriculum. Backers of PSSC physics, a "new" curriculum, suggest that its use fosters the development of more complex cognitive abilities in students than do more conventional curricula. The work done by Thompson and Schartz reports the CEEB Physics Achievement Test to be an appropriate criterion measure for both the content of PSSC and non-PSSC curricula. This test includes items requiring students to perform at the Knowledge, Comprehension, Application and Analysis levels of Bloom's Taxonomy of Educational Objectives-Cognitive Domain. The students' responses to these items, grouped and scored by levels of cognitive performance, should provide criteria for characterizing the similarities and differences between the PSSC and non-PSSC students.

Research Design and Procedure

This study was designed to test the following hypotheses:

a. Relationships between academic aptitude and performance at the Knowledge, Comprehension, Application and Analysis cognitive levels are not dependent on the type of high school physics background.

b. Students with non-PSSC physics background will score higher on the Knowledge level criterion than will the PSSC students.

c. Students with PSSC physics background will score higher on the Application and Analysis level criteria than will non-PSSC students.

Two samples of 370 each were drawn from a list of physics students who had completed Form LAC 2 of the CEEB Physics Achievement Test. One sample was of students with a PSSC physics background and the other sample was of students with a non-PSSC physics background. These samples were further reduced by the availability of Verbal and Mathematics SAT scores. The resulting samples were PSSC background = 369 and non-PSSC background = 359.

The CEEB Physics Achievement Test items were independently classified by two individuals into the following performance levels: Knowledge - 9 items, Comprehension - 17 items, Application - 26 items, and Analysis - 23 items. The interrater agreement was 66 percent. Reliabilities for these subscales were not reported.

Correlations between the SAT standardized scores and each of the cognitive level subscales were computed for each sample. Tests for group differences in Fisher Z transformed correlations were made.

Analysis of covariance as developed by Wilks and Gulliksen was used to test for difference in cognitive performance at each of the four levels. The SAT score was used as the covariate.

Findings

a. The PSSC sample and the non-PSSC sample were not equivalent in terms of the Verbal and Mathematics SAT scores. The differences favored the PSSC group at the 0.01 level.

b. The SAT scores correlated positively with performance on the cognitive level criteria. These relationships did not differ significantly for PSSC and non-PSSC groups.

c. The analysis of covariance produced the following results:

   1. The cognitive level criterion, Comprehension, did not satisfy the assumption of homogeneity of variance at the 0.01 level.

   2. The cognitive level criteria, Application and Analysis, did not satisfy the assumption of homogeneity of regression at the 0.01 level.

   3. The cognitive level criterion, Knowledge, resulted in a difference favoring the non-PSSC group.

   4. The cognitive level criterion, Analysis, resulted in a difference favoring the PSSC group.

Interpretations

Scholastic ability is equally important as a determinant of performance at the various cognitive levels for both PSSC and non-PSSC students. The investigator indicates that if the PSSC materials foster the development of more complex cognitive abilities than do the non-PSSC materials then the correlations of SAT scores with the cognitive measures should be less in the PSSC group than in the non-PSSC group.

The PSSC students performed significantly better on the higher level cognitive process measures. The investigator states that since this finding would be expected on the basis of the stated philosophy of the developers of the PSSC curriculum, it could be concluded that the "new" physics curriculum was more effective in developing the higher cognitive skills than were the more traditional curricula. Although these differences were significant, the adjusted means differed by less than one point. The investigator concludes that both the PSSC and non-PSSC curricula tend to develop the higher level cognitive skills.

Abstractor's Notes

It is difficult to know how much faith to place in the measures of cognitive level performance used in this study. The reliabilities of the subscales created from the CEEB Physics Achievement Test were not reported. These subscales ranged from 9 items to 26 items in length. The raw score means indicated that the average scores ranged from 26 percent up to 51 percent correct. Neither group did very well compared to the performance possible on the criteria. The difficulties that were experienced in meeting the assumptions for the analysis of covariance also may mean that the criteria should be examined more carefully. The adjusted means for these scales were not included in the report. It would be useful to have them since the differences were in terms of the adjusted means.

The conclusion that PSSC instructional materials were more effective in developing higher cognitive process skills is one way of accounting for the results reported in this study. The validity of this conclusion depends on a number of other conditions.

It was shown in the study that the samples of PSSC students and non-PSSC students were different in terms of SAT scores. This suggests that the populations represented by these samples were not equivalent for at least one characteristic and, most likely, there were other differences not detected by the investigator. The higher cognitive processes such as Application and Analysis correlated about 0.60 with SAT scores and accounted for 36 percent of the variance. The remaining 64 percent was assumed to be a measure of the performance at the specific cognitive level of interest. The investigator had to assume that no other initial differences in the groups were responsible for the results of the analysis of covariance. Since the PSSC group had higher SAT scores it is probable that they had other cognitive performance abilities to start with which were independent of those measured by the SAT. In short, the differences in the scores of the PSSC and non-PSSC groups on the Analysis cognitive level may have been there before the students took physics.

The types of communities, schools, and teachers which select and use PSSC materials may have more to do with the results than did the curriculum used. If the kinds of skills supported by the writers of

PSSC materials are valued and considered relevant by the teachers and others who make curriculum decisions then they are apt to select the PSSC materials for use. Students in schools where PSSC materials are selected probably already possess a higher degree of skills related to performance at the higher cognitive levels.

This study is a post-hoc analysis and the results should be treated as such. The investigator has concluded that "the results essentially supported the position of new curriculum writers that the PSSC instructional materials were more effective in developing higher cognitive process skills." There are other alternative claims and explanations which can be supported by evidence in this study. The data in this study are correlational and the causal inferences drawn are unwarranted.

EJ 088 494

Wolfson, Morton L., "A Consideration of Direct and Indirect Teaching Styles With Respect to Achievement and Retention of Learning in Science Classes." Journal of Research in Science Teaching, Vol. 10, No. 4:285-290, 1973.
Descriptors--*Academic Achievement, *Interaction Process Analysis, *Retention, *Teacher Characteristics, *Teaching Styles, Chemistry, Educational Research, Science Education, Secondary School Science, (Research Reports)

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Ronald D. Anderson and Kenneth D. Hopkins, University of Colorado.

Purpose

This study was undertaken to compare the Flanders I/D ratio of selected general science and chemistry teachers to their students achievement and learning retention.

Rationale

Studying teacher characteristics or the teacher's behaviors in the classroom as a means of learning what will ensure student learning is inadequate. The teacher-student interactions in the classroom must be examined. In addition to simple achievement, retention must be included as an indicator of student learning. Research is needed which relates patterns of teacher-student interaction to the student's short and long term learning.

Research Design and Procedure

Eight high school chemistry classes and six junior high school general science classes from one Suffolk County New York public school system were included in the study. No basis for the selection of these particular classes is given. The study is correlational in nature with no experimental treatment involved. The Flanders system was employed to code the classroom interaction twelve times during the year in general science and nine times in chemistry. In all classes an achievement test was given at the end of the school year in June and the same test was given to the students again four months later. The achievement test used in the chemistry classes was the New York State Regents examination and the one used in the general science classes was the College Entrance Book Company examination. No information on the reliability of the tests is provided. The scores from the first test were used in the analysis along with retention scores which were based on ratios of the retest score to the first test score for each class.

Apparently the fourteen classes involved in the study were taught by eight different teachers. For purposes of analysis it appears that the scores from all of the classes taught by a given teacher were consolidated and the analysis was done by teacher rather than by class.

Analysis of variance was used to determine if the achievement scores of the students taught by the various teachers differed significantly from each other. The analysis was repeated using the retention scores. The Duncan New Multiple Range Test was used to compare means after the F ratio was found to be significant in both analyses.

Further treatment of the data was based on the apparent post hoc finding that two of the chemistry teachers had rather high I/D ratios and two had rather low I/D ratios. The same pattern of two high and two low I/D ratios was found for the general science teachers.

Findings

Use of the Duncan New Multiple Range Test for comparison of means showed that in all four of the comparisons of a high I/D teacher with another high I/D teacher and of a low I/D teacher with another low I/D teacher, there was no significant difference in the means on the achievement test. In the case of the retention scores three out of four such comparisons of low I/D teachers with each other or of high I/D teachers with each other were not significant. When looking at the various comparisons of a low I/D teacher with a high I/D teacher, six out of eight such comparisons showed significantly different means on the achievement test and eight out of eight such comparisons showed significantly different means on the retention score. The direction of these differences was in favor of higher student scores for the teachers with higher I/D ratios.

Interpretations

The author concludes that in chemistry or general science, students will achieve higher scores on an achievement test and on a retention test if they are taught by a high I/D teacher. He infers that the "flexibility" of the high I/D teachers (their I/D ratios were found to vary more from one time to another than those of the low I/D teachers) may account for the differences in achievement and retention of the students. Since greater differences had been found in the retention means than in the achievement means he speculates that the retention test may be more sensitive to differences and thus the retention test is perhaps a better indicator of learning.

Abstractor's Notes

This report provides an example of research in which the investigator had a viable purpose and collected usable data yet failed to publish an adequate study due to faulty data analysis and interpretation. It has several serious defects.

The study is improperly interpreted; cause-and-effect conclusions are drawn where only an association is warranted, viz., (p. 289). "Furthermore we can conclude that students of teachers with a lower I/D ratio." But could it not be the other way around -- higher achieving students and classes cause teachers to be more encouraging in their responses to the class? Wolfson violates one of the most elemental research axioms, correlation does not necessarily mean causation.

A related problem results from the study's failure to rule out compelling selection threats to the internal validity. Almost certainly the classes were not equal in IQ, socio-economic status and many other variables related to science achievement. To disregard all but the I/D class differences in explaining achievement differences among the classes is not justifiable. The study could have increased its credibility greatly had one or two meaningful variables, such as IQ or standardized achievement score, been used as covariates.

A third serious weakness results from an improper application of ANOVA. Classes are a random factor, i.e., the investigator desires to discover a finding generalizable to other classes like the ones in the study. The fixed effect of interest is the I/D level. The analysis is oblivious to this fundamental distinction and hence does not directly answer the researcher's question. But what are the consequences? The analysis should have reflected the implicit hierarchical design -- the I/D classification (high or low), the class factor nested within I/D, and pupils nested within class. The correct analysis would use the mean square for classes within I/D level as the proper denominator for the F-test comparing the means of the high and low I/D groups. This F-ratio would require an enormous critical F-value since there would only be 2 degrees of freedom in the error term. If there were no significant differences between classes within I/D category, the SS and df from classes within I/D level and pupils within class could be pooled to yield an error term with more power. An "eyeball examination" of the means suggests that this latter situation probably pertains.

Wolfson performed a one factor ANOVA comparing the four classes, which confounds I/D with class effects. He then proceeded to make multiple comparisons among classes to "back into" inferences regarding high-low I/D differences. Multiple comparisons are never appropriate with a random factor such as classes, but multiple comparison would have been unnecessary had the proper hierarchical analysis been employed since only two means for the I/D effect would have been involved.

Parenthetically it can be noted that the Duncan Multiple comparison technique is a poor choice even if multiple comparisons had been in order. With the Duncan technique α varies depending on the number of means in the set being examined. With four means with an apparent α of .05, the actual α is .14 when comparing the most extreme means. Rarely does a researcher want to reject some mean differences at a much more relaxed type-I error probability than others -- at least if he does, he wants to control α, something he cannot do with the Duncan technique.

Several minor problems deserve mention. No information is given as to how the total number of students was reduced to 160 or why there were yet 168 in the ANOVA (Table III). How many cases were lost between the posttest and delayed posttest four months later? The sample is inadequately described which limits the study of external validity; no information is given regarding ethnicity, socio-economic status, IQ, or sex. Impressionistic information is considerably better than no data at all.

The manner in which the retention test was handled leaves much to be desired. The ratio used allows regression effects to confound the dependent variable and hence reduces its reliability. Covarying the end-of-year test or using a repeated measures analysis would have been preferable.

The investigator could have increased power and obtained direct information on whether science content (chemistry or general science) interacts with the I/D factor by converting scores on both tests to standard scores and employing all observations in the same analysis. The three factor design then would be science content by I/D level, with classes nested within I/D (and pupils nested within the three factors). Examination of the interaction of I/D level by science content would indicate whether or not the difference in high and low I/D levels was generalizable across chemistry and general science.