DOCUMENT RESUME

ED 132 000 SE 021 331

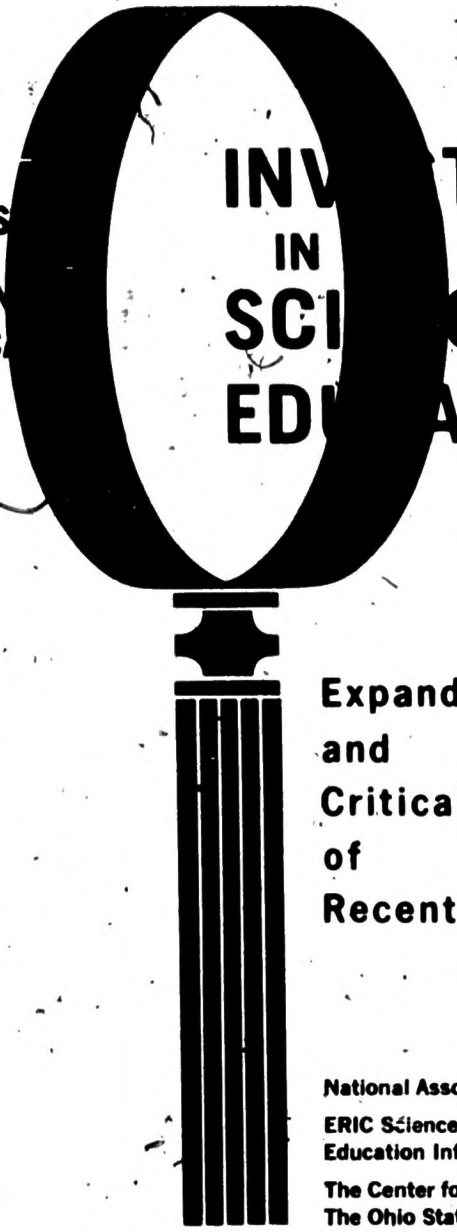
AUTHOR Helgeson, Stanley L., Ed.; Blosser, Patricia E., Ed.
TITLE Investigations in Science Education, Vol. 2 No. 1. Expanded Abstracts and Critical Analyses of Recent Research.
INSTITUTION Ohio State Univ., Columbus. Center for Science and Mathematics Education.
PUB DATE 76
NOTE 66p.
AVAILABLE FROM Information Reference Center (ERIC/IRC), The Ohio State University, 1200 Chambers Road, 3rd Floor, Columbus, Ohio 43212 (Subscription $6.00, $1.75 single copy)

EDRS PRICE MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS *Abstracts; Cognitive Development; *Educational Research; *Evaluation; Instruction; *Literature Reviews; Research; *Science Education; State of the Art Reviews
IDENTIFIERS Research Reviews

ABSTRACT

This volume contains analyses of 12 recent studies in science education. The research topics of the studies include cognitive development, instruction, statistical methodology, teacher education, and testing. Each analysis is written by a science education specialist and contains an expanded abstract identifying the purpose, rationale, methodology, findings, and conclusions of the study. In addition, the analyst provides a critical evaluation of the research and recommendations for further study. (MH)

Vol. 2, No. 1 1976



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 Center for Science, Mathematics,
and Environmental Education

Published Quarterly by

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

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

INDIVIDUAL ANALYSES

Vol. 2, No. 1, 1976

NOTES

from the Editor

This volume contains analyses of research studies in three categories. The first group, INDIVIDUAL ANALYSES, continues the practice, where appropriate, of critiquing studies on an individual basis. The other two groups, COGNITIVE DEVELOPMENT and INSTRUCTION, contain analyses of studies related to the themes reflected by the category titles. These "clusters" of reviews are an attempt to present research reports with some common basis to better represent the current state of the art in science education research.

Publishable responses to the analyses are encouraged as well as suggestions for improving Investigations in Science Education.

Stanley L. Helgeson
Editor

Patricia E. Blosser
Associate Editor

EJ 034 550

Carus, Frank E. and Lawrence F. Lowery, "The Usefulness of Closed Circuit Television and Psycho-Galvanic Response in the Training of Student Teachers." Journal of Research in Science Teaching, Vol. 7, No. 4:353-361, 1970.

Descriptors--*Evaluation Techniques, Evaluation Methods, Preservice Education, *Physiology, *Student Teaching, Science Education, *Teacher Behavior, Teaching Experience, *Video Tape Recordings

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Michael Szabo, The Pennsylvania State University.

Purpose

The purpose of this study is loosely defined as to test the usefulness of videotapes in teachers' self-evaluation and explore "the physiological effects of such self-evaluation." After reading the study the reviewer infers that Carus and Lowery were hypothesizing that classroom behaviors of student teachers would change as a function of having a videotape transcripts and/or having galvanic skin response monitored while they rated their own teaching behaviors. The criteria of change were 16 categories of classroom occurrences as interpreted by the OScAR instrument. Mean gain scores on these 16 categories were computed from observations made early (first week) and late (8th week) during the student teaching experience.

Rationale

Little or no attempt to provide a theoretical or practical model to be tested or to guide the study was presented in the article. The reviewer again takes the liberty to infer that the major rationale for the study was "No research has been done in the area of galvanic skin response." Some literature is cited regarding the efficacy of using videotapes in improving teachers' critical analysis ability. Although a routine test for the interaction of galvanic skin response (PGR) and videotape was performed in the analysis of variance, no rationale for such an interaction was developed.

This study assumes that there is a relationship between critical self-evaluation analysis of teaching behavior and positive changes in either student attitude or achievement. It also assumes that individual student differences are unimportant regarding the subject matter and the amount of learning that takes place.

Design and Procedures

The design employed in this study was a 2 factor (2 levels each) design with pre- and post-tests of external observations. The two independent variates included two levels of videotape use (videotape

available or not available during self-coding) and two levels of PGR monitoring (present or absent) during the self-evaluation. The criterion variate was the mean gain pre-post-test observation using an author-modified version of OScAR. Ratings were completed by the researchers. The sample was composed of 28 science teachers who, during their student teaching experience, were randomly assigned to (one of) four treatment groups. Presumably there were seven subjects in each of the cells indicated in the Figure 1 below.

	Videotape	No Videotape
PGR	Condition 1 CRITERIA	Condition 2 = MEAN
No PGR	Condition 3 GAIN SCORES (OScAR)	Condition 4

Figure 1. Cell Design

During the student teaching experience, each subject was assigned identical lessons to teach during three observational periods of one half-hour each. At some later time, each student rated his performance during these periods using the 16 categories of classroom occurrence in OScAR. Student self-ratings were not used as part of the criterion variates. Under condition 1, ratings were completed with the polygraph attached while the student viewed the videotape. Condition 2 included both the polygraph and videotapes of student performance. Condition 3 used the videotape but no polygraph, and condition 4 employed neither videotape nor polygraph. It was not stated whether students had access to their physiological response data or interpretations of response patterns during the experiment.

Data were analyzed using a 2 x 2 analysis of variance model with the criterion of mean gain score on the OScAR observations. One analysis of variance was tabulated for each of the 16 categories in the modified OScAR. In addition, patterns of correlation between polygraph readings and each of the 16 category scores for student ratings at weeks 4, 6, and 8 were presented for condition 1 and 2.

Findings

The analyses of variance revealed no significant main polygraph effects and no significant interaction effects for any of the 16 criterion variates. The videotape main effect was significant for the following four OScAR categories: Teacher Introduces a New Activity, Evidence of Preparation, Individualization of Instruction, and Teacher Uses Distributive Material. No indication was given

whether the significant gain in mean scores favored the videotape or the no videotape group. Regarding the correlations obtained from subjects in condition 1, seven of the 48 correlations were significant at the .05 level. Under condition 2, only three correlations were significantly different from zero.

Interpretation

The authors claim, without elaboration or support of evidence, that the polygraph used in conjunction with videotape "proved to be an important tool in isolating covert emotional reactions of student teachers to their teaching behavior" (p. 357). They further state, without elaboration, that "The PGR readings, when used in conjunction with the videotapes, increased the effectiveness of evaluation by revealing an entirely new set of variables that influence teaching behaviors" (p. 359).

ABSTRACTOR'S ANALYSIS

The analysis of the report is divided into five subcategories which deal with design, measurement, analysis, interpretation, and clarity and conceptual base of the report.

Design

The design was well conceived with random assignment of subjects to (one of) four treatment conditions. Random assignment is one of the assumptions under the use of analysis of variance. The chief problem in this study is the sample size of seven subjects per cell which severely limits the interpretations one can make. The likelihood that the results are attributable to chance increases as the sample size decreases. That is, the power of the test of significance is severely jeopardized. This factor is seen as a severe limitation to the interpretation and usability of the results generated by Carus and Lowery. They comment about the most significant change in teaching behavior: "Despite the small sample size... ." It is more likely that the results cannot be reliably interpreted because of small sample size.

Measurement

The authors are to be congratulated on the selection and use of a tested and viable observational instrument for the criterion variate. The reviewer, however, raises the question that neither a description nor a rationale for the modification of OScAR was presented. It is possible that any modification can change the nature of the instrument. Whenever potential changes are noted, a new assessment of the reliability and validity of the instrument must be made to eliminate the hypothesis that the results may be attributable to chance measurement errors. Furthermore, the assessment should be conducted prior to the

study in which final data are to be gathered. The basic minimum is a reliability estimate obtained on the sample used in the study. No reliability figures were reported for either the observers' pre-test and post-test observational ratings or the reliability of the students' self-ratings for any of the categories of OScAR. If the former are unreliable, it is clear the findings cannot be trusted. If the latter are unreliable, the unfortunate practice of having prospective science teachers make judgments on data which are suspect might again be perpetrated.

The authors have attempted to document the validity of the modified OScAR, but fail to tell us the type of validity, the criteria for acceptable validity, and whether the data met the criteria. We must conclude from this that the reliability and validity of the chief instrument in this study are open to suspicion.

A second point regarding the reliability of the pre- and post-test data is in order. Medley and Mitzel (2) have shown that there are substantial day-to-day variations in observed teacher performances. Given these variations, repeated measurement by a single observer holds the most potential for obtaining sufficiently reliable observations. The data presented by Medley and Mitzel suggest that as many as 16 different observations (of 15 minutes each) by a single observer are needed to achieve a reliability coefficient of .90. The authors (apparently) have relied upon two 30 minute segments to gather pre- and post-test data. It is likely that insufficient observational time was allowed to gather data not suceptible to day-to-day variations in teaching performance. It appears that the methodology of measuring the criterion variate in this study is rather suspect.

Analysis

The authors chose to use as their criterion variate the gain between pre-test scores and post-test scores based upon observational data using the modified OScAR instrument. The analysis of gain scores is quite complex since even if the pre- and post-test observations are themselves reliable, their difference can be quite unreliable [Cronbach and Furby (1)]. Carus and Lowery used the acceptable procedure of analyzing mean gain score rather than raw gain scores. An acceptable alternative open to this procedure was to use a repeated measures analysis of variance design in which the pre- and post-test data are the repeated measurements.

An additional reason to use the repeated measures statistic is the small sample size. When the repeated measures are substantially correlated as they are in a pre-post-test design and the sample is small, the power of a repeated measure design is considerably greater than the power of a non-repeated measures design.

The authors present tables showing the eight-week patterns of correlations between PGR and OScAR data. Of the 96 correlations observed, ten were significantly different from zero ($p < .05$). Since five of these correlations can be expected to be significant

by chance alone, a problem arises when one attempts to distinguish between the significant and nonsignificant correlations.

Whenever multiple criterion variates are analyzed, the question of independence (low intercorrelations) among the variates is of concern. One practical problem associated with high intercorrelations among the criteria surfaces is erroneously drawing different conclusions about the same phenomena. If, for example, the two categories from OScAR entitled "Evidence of Preparation" and "Individualization of Instruction" were uncorrelated, two separate conclusions might be justified. On the other hand, a high correlation between these two categories might imply that these two categories measure the same concept or behavior. It follows that two separate conclusions would not be warranted.

Two alternative solutions to this problem are available. The first is to generate the intercorrelation matrix of 16 categories that were used as the criterion variates and inspect them for significant correlations. The second solution is to use a statistical treatment which accounts for intercorrelated criterion variates such as multivariate analysis of variance design.

Clarity and Conceptual Base of the Report.

The reviewer found a number of disconcerting omissions and was forced to make some assumptions in evaluating this research study. For example, it was not immediately clear whether pre-service or in-service science teachers comprised the sample. In addition, no mention was made of any training in the use of OScAR by either the students who participated in the study or by the observers. The lack of information on the reliability and validity of the modifications of OScAR has already been mentioned. A related observation is that since the researchers who coded teacher behavior (criterion) also knew which subjects were assigned to which treatments, the reviewer raises the issue that such observations may have been biased (Experimenter Bias Effect).

The final point of clarity involves the authors' conclusions regarding the usefulness of PGR in revealing an entirely new set of variables that influence teaching behavior. Perhaps the authors would share this information (supportive evidence) with us in the future.

Two conceptual problems keep coming to the reviewer's mind throughout the reading of this study. No rationale for any expected relation between the criterion of classroom behavior and the use of PGR was presented. A rationale for interaction between the use of PGR and videotape was not developed. It is likely that examination of the literature of PGR-related research from psychology and from educational psychology would provide a minimum theoretical base for the study. The lack of rationale (and resulting hypotheses to guide the study) permits the authors to minimize any discussion of the statistically significant findings and to expand the discussion of results which are not supported by data presented in the study.

A more severe problem is echoed in the question, "Why train teachers to focus on improvement of classroom behaviors?" Until research establishes at the minimum a strong link between teacher behavior and student attitude or student achievement, one must seriously consider this question.

Summary

The authors present a research study which partially supports the usefulness of videotapes of teaching performance in influencing classroom behavior but which does not support the use of PGR relative to this criterion. A more reasonable conclusion is to defer judgment about these variables until a more tightly designed study can be conducted.

REFERENCES

1. Cronbach, L. J. and Lita Furby, "How Should We Measure 'Change' - Or Should We?" Psychological Bulletin, (74):68-80, 1970.

2. Medley, D. M. and H. E. Mitzel, "Application of Analysis of Variance to the Estimation of the Reliability of Observations of Teachers' Classroom Behavior." Journal of Experimental Education, (27):23-35, 1958.

EJ 080 456

Grobe, Cary H., "A Regression Approach to Evaluating Instructional Programs in Science." Journal of Research in Science Teaching, Vol. 10, No. 1:55-62, 1973.

Descriptors--*Multiple Regression Analysis, *Research Methodology, *Science Education, *Statistical Analysis, *Teaching Procedures, Teaching Techniques

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

Purpose

The basic purpose of this paper was not to report the findings of a scientific study, but to present the multiple regression approach for use in the evaluation of instructional strategies in science. In order to illustrate this methodological approach, data from a previous study by the author involving two methods of college biology instruction (audio-tutorial and conventional) were analyzed.

Rationale

Many approaches to evaluating instructional programs are feasible. The author presents multiple linear regression as a powerful and simple approach to the resolution of research problems. The author points out the advantages of this approach over the more commonly used analytical approaches such as analysis of variance.

Research Design

The study to which multiple linear regression was applied involved 38 subjects who were assigned randomly to receive the A-T college biology course and 41 subjects assigned randomly to a conventional biology course. A pre-test - post-test control group design was used and subjects were classified into one of three aptitude levels according to composite scores on the American College Test. The design was a treatment by levels with two treatments (A-T and conventional biology instruction) and three levels of aptitude within each treatment (high, middle, and low). In order to evaluate student achievement in biology, Section I of the College Entrance Examination Board Advanced Placement Exam in Biology (CEEB) was administered. No further description is given of the study by the author because the report focuses on the methodology used to analyze results rather than outcome of the study itself.

Application of the Methodology

Four questions in the form of linear models amenable to solution by multiple linear regression are presented. The discussion of mathematical details and computational operations is deliberately

deleted in the report. Instead, emphasis is placed on data interpretation.

Question 1. "Was there a significant difference in CEEB scores between subjects who had received A-T and conventional biology courses?"

The linear model

$$Y = a_0U_0 + a_1X_1 + a_2X_2 + e$$

Which reduced to

$$Y = a_0U_0 + a_1X_1 + e$$

was formulated to represent question 1 where

Y = CEEB scores, the criterion vector

X_1 = 1 score if vector Y was from a member of the A-T group; and 0 if otherwise.

X_2 = 1 if score in vector Y was from a member of the conventional group; and 0 if otherwise

A_0 = a unit vector (all 1's generated by the computer program)

A_1 = the regression coefficient for X_1

A_2 = the regression coefficient for X_2

e = the error associated with the prediction system

A sample of the coding of punch cards for the regression solution of question 1 follows and the regression solution of the equation given.

It was found that a multiple correlation coefficient (R) of 0.075 resulted from the above analysis. The interpretation of R was the same as that of the Pearson product-moment correlation coefficient (r) (1.00 being a perfect positive correlation and - 1.00 being a perfect negative correlation). The author concludes that there was no significant difference between being in the A-T group or being in the conventional group on the CEEB. Other questions and multiple linear regression equations which would give a solution to the question are presented in the article. Because of limitations of space in this review only the questions and their solutions will be given.

Question 2. "Were there significant differences in CEEB scores between subjects of similar aptitudes who had received A-T and conventional biology courses?"

The R value found by the regression solution to Question 2 was 0.309 indicating no significant difference in biology achievement at each aptitude level. R^2 (.0954) indicated that 9.5 percent of the CEEB variance was accountable from a knowledge of the subject aptitude and group membership.

Question 3. "Were there significant differences in CEEB scores within A-T and conventional groups associated with the subject aptitudes?"

A linear equation was formulated. Because of its similarity to that formulated in Question 1, it was unnecessary to solve the equation via computer analysis. Instead Snedecor's F ratio was employed. The F value of 1.69 indicated that there was no significant difference at the 0.05 level in CEEB scores between levels of aptitude after allowing for the influence of instructional methods.

Question 4. "Were the differences in CEEB scores the same for all three aptitude levels within the A-T and conventional groups?"

Solution of the linear equation for analyses of the above question gave an R^2 value of 0.087. This R^2 value was compared to the R^2 obtained for question 2 (R^2 = 0.095) using Snedecor's F ratio. An F value of 0.352 indicated that the differences between CEEB scores for pairs of aptitude levels were constant within the A-T and conventional group, that is, there was no interaction effect between aptitude levels and instructional methods.

Indications

The use of the multiple linear regression method indicated that 9.5 percent of CEEB variance accounted for the students' aptitude and instructional method. This method of analyzing data enables the researcher to determine the percentage of the variance due to the factors being analyzed.

The author concludes by citing references that would be useful to the reader interested in learning more about the application of multiple regression to experimental design.

ABSTRACTOR'S ANALYSIS

This research report offers a valid alternative method of analyzing research data. The author attempts to present the method in a nonmathematical way to enable the reader to understand the approach without becoming too heavily involved in the computations and theory underlying it.

The author's intent is to compare this method with the analysis of variance. In so doing, limitations of analysis of variance are noted. It is stated for the ANOVA, data must be arranged in standard designs, i.e., matched pairs of equal cell frequencies. These limitations noted by the author are not in accord with those noted by Winer (2) and other statisticians. Randomly assigning subjects to treatments is considered preferable over using matched pairs (1) and there are definite ways of handling unequal cell frequencies.

There are several ways in which this research report could have presented the linear multiple regression approach more clearly.

1. The author uses data from a study in which no significant differences were found. It would have been more interesting to use data which, when analyzed, would have produced significant differences.

2. The questions asked are not framed in such a way that a direct comparison can be made with questions analyzed using an analysis of variance. What are the main effects? What is the interaction effect? It would have been most helpful if the data were analyzed using both methods and comparisons shown.

3. In the description of the study, a pre-test - post-test control group design was employed. How is this analyzed using this method? It is not clear from the questions posed if the data were analyzed.

4. A more careful use of symbols would have made the equations easier to follow. For example, "U" is not defined. A small "c" appears in the equation but is coded as a capital "C".

Although this report may have the shortcomings listed above, the author shows quite clearly how regression equations can be used in data analysis. Because of limitations of space it is sometimes very difficult to condense a statistical method in such a way to make it comprehensible.

REFERENCES

1. Travers, R. M. An Introduction to Educational Research. New York: McGraw-Hill Book Company, 1971.

2. Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill Book Company, 1971.

17

EJ 049 058

Hoch, Loren L., "Attitude Change as a Result of Sex Education." Journal of Research in Science Teaching, Vol. 8, No. 4:363-367, 1971.

Descriptors--*Attitudes, Biology, *Changing Attitudes, Evaluation, Instruction, Knowledge Level, *Sex Education, Secondary School Science

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

Purpose

The author attempted to discover whether or not a nonjudgmental sex education unit taught in a high school biology class significantly dispelled ignorance of sexual attitudes and facts.

Rationale

The research was conducted within the confines of four sections of high school biology taught by the author in a large northern Indiana high school during a two-week period. The Institute for Sex Research at Indiana University cooperated with the study.

Research Design and Procedure

One hundred students, 50 in the control group and 50 in the experimental group, participated in the study. Ten class periods were included, each 50 minutes long. The six dependent variables were:

variable 1: factual knowledge;

variable 2: male permissiveness;

variable 3: female permissiveness;

variable 4: confidence in ability to make proper decisions in the future involving their own sexual behavior (anxiety level);

variable 5: attitude toward population control, family planning, and birth control; and

variable 6: attitude toward sexual deviates.

Three weeks before the two-week unit of instruction began, pre-study tests for each of the six variables were given to the experimental and control groups. Post-instruction tests were administered when the unit was completed.

Variable 1 was measured by a 58 item terminology test adapted from "Sex and the Whole Person" (2) and from a questionnaire by M. Neil Macintyre, Department of Anatomy, Case-Western Reserve University. An attitude scale by Ira L. Reiss, University of Minnesota,(1) was used to measure variables 2 and 3. The author developed items, validated by a panel of experts, for measuring variables 4 through 6.

A t-test was applied to determine pre and post study results, summarized in eight tables. Tables IX through XIV present an analysis of variance for each of the 6 variables.

Findings

The findings of this study show the following:

1. Factual knowledge is increased as a result of sex education.
2. Students do not become more permissive in their attitudes involving sexual behavior as a result of sex education.
3. Students become more confident in their ability to make decisions in their own sexual behavior.
4. Students become more aware and more liberal in their views about population control, family planning, birth control and abortion.
5. Students are less hostile and more understanding of sexual deviates as a result of sex education.
6. Much of the success of sex education depends on the instructor.

Interpretations

The author concludes that a major objective of the unit, that of dispelling sexual ignorance, was partially met. Attitudinal changes concerning permissiveness for males and females were not shown at a significant level, seemingly negating one criticism of sex education: that moral standards of participants are lowered. Significant change in students' confidence in their ability to make good future decisions concerning sexual matters was shown. A trend toward more liberal thinking about population control, family planning, birth control and abortion developed during the study. Sex education brought about more accepting attitudes toward sexual deviates. The author stresses the value of keeping a sense of humor during class discussion as well as avoiding giving an "I know all the answers" impression to students while conducting sex education programs.

ABSTRACTOR'S ANALYSIS

With the subject of sex education becoming more important in the secondary curriculum it is necessary to have studies such as this one to collect data on the subject.

The results are not surprising, but they do add to the support of having sex education in the schools. Neither the experimental model nor the treatment of the data were unique; only the topic is of importance.

It would be interesting to conduct the same study in different regions of the country and compare the results. The attitudinal portion of the study is probably of greatest importance.

REFERENCES

1. Reiss, Ira L. The Social Context of Premarital Sexual Permissiveness. New York: Holt, Reinholt and Winston, 1967.

2. Resource Packet, Southern California-Arizona Conference, the Methodist Church, 1966.

EJ 022 863

Walbesser, Henry H. and Heather L. Carter, "The Effect on Test Results of Changes in Task and Response Format Required by Altering the Test Administration from an Individual to a Group Form." Journal of Research in Science Teaching, Vol. 7, No. 1:1-8, 1970.
Descriptors--*Academic Achievement, *Evaluation, *Elementary School Science, *Measurement Instruments, *Student Testing, Tests, Test Validity, Test Reliability

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

Purpose

The purpose of this study was to assess the test administration format of competency measures for Parts E and F of SAPA I (Science-A Process Approach I). The hypothesis tested was:

> Changing the assessment task from an individual to a group format decreases the number of subjects able to complete the task.

This research hypothesis may be altered into the following probabilistic hypothesis:

> The difference in performance results for a given objective within a given process is as likely to favor the group (G) competency measure data as it is the individual (I) competency measure data. Pr (I - G = 0) = 0.5.

Rationale

In the development of SAPA I by The Commission on Science Education of the American Association for the Advancement of Science, a commitment was made to systematic evaluation. This report deals with individual student performance at the end of each instructional exercise. Each of the exercises is organized around one of the thirteen processes identified by AAAS. The hierarchy of learning is based on the psychological premises of Robert Gagné. Reference is made to construction and validation procedures for the process hierarchies.

In the developmental use of the SAPA I materials, many teachers expressed concern over the amount of time required to administer competency measures to each individual student at the completion of each of approximately 20 exercises each school year. Because of the concern of these teachers as to the use of instructional time, individual and group competency measures were constructed for each exercise in Parts E and F of SAPA I.

Research Design and Procedure

A set of nine performance classes were constructed. Each class was identified by an action verb: naming, identifying, ordering, constructing, describing, demonstrating, stating a rule, applying a rule, and distinguishing. Outlined in the report are examples of task format and response format for each of the nine performance classes for both individual and group competency measures. Two detailed examples of individual and group tasks are given for the performance class demonstrating. Generally the tester can directly observe each learner's attempt to exhibit a desired behavior in the individual format. In the group situation the tester must rely on paper and pencil recordings of the learner's performance on a task. A student may select illustrations from a set of illustrations and order them according to a procedural step specified in the task. Several students can complete this task simultaneously.

The manipulated variable for the study was the task and response format. The types of task and response formats used are outlined in the report. The differences necessitated by the individual or group size are noted. For example, in ordering in the individual format, the student arranged a given set of materials. In the group format, the pupil placed numerals on a set of illustrations to indicate a sequence.

The responding variable was defined as the direction of the difference between the individual (I) and group (G) competency measure results for each objective of each exercise in Parts E and F of the SAPA I materials being developed. This binomial variable was positive if I mean minus G mean was positive.

The G and I competency measure data were gathered from random samples of children in tryout classes in each of 14 test centers in the United States in 1966-68. The number of children tested is not reported.

The teachers were randomly assigned to one of two test administration schedules. One schedule was G, then I, then G, alternating the form for all exercises taught. The other schedule was I, then G, then I, etc. The test results for all teachers were pooled. Each teacher reported I and G results obtained from a random sample of children in each class. No comparisons were made if the number of subjects tested in a pair of G and I competency measures differed by more than two individuals.

The design of the study cannot be directly described by Campbell and Stanley nomenclature. Subjects were "randomly" assigned to a treatment that varied in sequence. Post-tests were administered. The design may be represented as R X O.

The data on the responding variable were organized by process. The processes are Controlling Variables, Defining Operationally, Formulating Hypotheses and Interpreting Data. Exact probabilities were computed and the decision of support or refutation was based on these results.

Findings

The hypothesis that Pr (I - G = 0) = 0.5 was rejected for each of the four processes. According to the investigator, the I results were consistently higher than the G results. A table is provided that gives the number of objectives and observed differences I > G for each of the four processes.

The data were also organized in terms of the nine defined performance classes with "naming" subdivided into oral and written components. The hypothesis of equally likely differences was accepted for naming-written, identifying, and stating a rule in writing.

Interpretations

The three performance classes, where Pr (I - G = 0) = 0.5 was accepted, have in common the response format of writing. The investigator concluded that the evaluation of behaviors related to these three classes can employ group tasks as substitutes for individually administered tasks.

Since this study was initiated as a result of teacher concern for the use of classroom time, the investigator concluded that direct performance assessment (I) does take more time but is a more suitable method of estimating performance than the more convenient group (G) format.

ABSTRACTOR'S ANALYSIS

In elementary science education, the SAPA program, science as process, and performance objectives, have become closely related terms. The delineation of the thirteen processes by AAAS has served as a useful organizational factor to assess student development in process as one of several components of scientific literacy.

The assessment of process, as with other components of science education, has two time facets. Short term assessments can be administered at convenient times during the instructional sequence. Long-range assessment involves use of the processes at a future time removed from the initial instruction. This study dealt only with the short term assessment.

According to the learning hierarchies patterned after the work of Gagné, the four integrated processes assessed in this study are dependent on the mastery of subordinate basic processes. Since assessment was made of levels E and F (grade levels 4 and 5), what were the experiences of the children with the processes to be mastered in parts A - D (grade levels K-3)? Did these children participate in the development of levels A - D? Were they competent in the use of the subordinate processes? If so, how were they assessed?

A fundamental issue that has continued to be an active factor in the implementation of SAPA is the operational definition of mastery. According to Gagné, a student must master each step in a learning hierarchy before he can progress to a future step. How many tasks were used to measure each objective in this study? Was mastery defined as seven out of eight tasks acceptable? Possibly five out of ten? Did students perform at a one-hundred percent competence level? An operational definition of how objectives were met is needed in order to interpret the stated results.

Several problems exist with the sample size. How many students were involved? How many teachers? How many classes? How were differences between classroom settings controlled? Did all students have an equal time exposure to SAPA as a total program? How much time did each student spend in science? With SAPA? Unless the instructional atmosphere is clarified and the sample size specified, it is impossible to generalize to a larger population.

The only data reported are the number of objectives per process or per performance class and observed differences I > G. Since no sample calculation is provided and no sample size is specified, the reader cannot follow the method used to report and interpret data. More detail is needed with a sample calculation of probability.

The method of alternating I and G measures is not explained adequately. Why did both groups of teachers use both forms? Would it have been possible to have one set of teachers use the I measures and another group use the G measures? The manipulated and responding variables would remain the same as in the original study. Perhaps matched pairs of students would be desirable?

How were teachers "randomly" assigned to a particular test schedule? How were children selected "randomly" from tryout classes? The entire problem of "who the sample is" seriously undermines the generalizability of this study.

Since all the students and teachers used in the study were in test centers where the SAPA I materials were being developed, the halo effect which "plagues" all trial development assessment may have been operational. Can the assumption be made that all subjects were affected equally and therefore this effect was not an uncontrolled variable?

Fundamental to curriculum assessment is the issue of continuity between objectives-instruction-evaluation. In the SAPA I materials, a major effort was made to develop this continuity in the written teacher guide booklets. The objectives are performance oriented, the instruction includes a high level of hands-on activities, and the evaluation tasks are keyed directly to the objectives. If the evaluation tasks (competency measures) are keyed to the objectives and the instruction mode is of a student participation nature, then it should follow that the evaluation is conducted in a task-oriented manner. Hands-on instruction should be assessed by hands-on evaluation techniques. Written instruction can be adequately assessed by written tasks.

An additional argument for hands-on evaluation stems from the child's cognitive development. It is questionable whether nine and ten year olds, for the most part, can verbalize and generalize many of their experiences. Written evaluations many times require the child to generalize and then verbalize. The construction of written tasks is very difficult because of this problem.

Even though logical arguments may be developed to support individual task oriented assessment keyed directly to performance objectives, this study alone does not support these arguments. Perhaps with additional controlled research, these hypotheses of the effectiveness of individual task oriented competency measures may be supported.

Since the time of this study, 1966-68, the SAPA I materials have been marketed and revised. It is interesting to note that in SAPA II, modules 61-105 for grade levels four-six, both group and individual competency measures are provided. Module 62, the first lesson on controlling variables, has five objectives that use the performance classes identifying and demonstrating. The student is assessed as to competence in demonstrating by the use of a paper and pencil device in a group setting!

COGNITIVE DEVELOPMENT

ED 065 325

Power, C. N., "The Effects of Pupil Involvement in Classroom Interaction of Science Achievement." In Research 1971, R. P. Tisher (ed.), Australian Science Education Research Association, Clayton, Victoria, Australia, 1971.

Descriptors--*Curriculum, *Evaluation, Educational Research, Interaction Process Analysis, Instruction, *Learning, *Science Education, Student Characteristics, *Teacher Education

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

Purpose

The stated purpose of the study described in this report was to explore the differences in science achievement between groups of students who illustrated different levels of classroom participation.

Rationale

In science, learning is widely viewed as an active process rather than a passive reception. If it is to be active, participation of the learner must then be a necessary condition for learning to occur. However, some research seems to suggest that these widely accepted assumptions are not true for all learners in all situations. Based on these research studies, it is essential that we know for what kinds of students active participation will help learning and for what kinds of students active participation will hinder this learning.

Research Design and Procedure

The subjects used in this study were eighth grade students. How many were used is not included in the report nor is any description of these provided for the reader. It is implied that there was some sort of treatment with no control groups. The study can be inferred to be a one-group pre-post design in which

$$Q_1 \quad X \quad Q_2$$

Q_1 = Pre-treatment videotape
Pretest of pupil achievement
Measures of
a. pupil cognitive abilities
b. pupil personality
c. pupil expectation of success in science
d. pupil achievement motivation
e. pupil sociometric status
f. pupil attitude toward science

are implied instructional experiences during which four videotapes were made - this time period was not specified.

Q_2 = Posttest of pupil achievement

The independent variables of the study were gain in achievement and "Involvement with teacher" as measured by the analysis of the videotape. The dependent variables of the study appear to be pupil characteristics as measured in the pretest measures. The measures for either the independent or dependent variables are not identified in the study.

Findings

Using a multiple discriminate analysis and three-way analyses of variance, three functions were found to be correlated with the combined independent variables of achievement gain and frequency of involvement with teacher. These functions were an achievement syndrome (pupils high in need expectation are seen to be involved more with teachers and gain in achievement), social orientation (student achievement is seen to be related to group dependence as reflected in their personality characteristics), and social maturity (a pupil's social economic status and role expectation seem to be related to involvement and achievement).

Interpretations

The findings of this study support the thesis that pupils do respond to instruction in quite different ways depending on a variety of personality characteristics. Thus it is essential that curricular materials be developed which provide for "greater pupil participation and involvement in the learning task."

ABSTRACTOR'S ANALYSIS

Knowing how much a pupil will gain from a science learning experience is a key question of very practical significance to every science teacher. The title of this study leads the reader to believe that here is an experimental study which will help in our understanding of the relationship between pupil involvement and pupil achievement in science. When the report is carefully read, this is not so. Rather what seems to be here is a technical series of explorational analyses in which many unspecified data were correlated to see what one will find. While the design of the study may be reasonably inferred, the author does not describe it so that the reader has a picture of what was done when and to whom. The variables in the study are named but their definition is neither given or are the means for their measurement described. The introduction would lead you to expect "active participation" to be one of the independent variables. Later it appears that some of the pupil data were grouped based on "frequent," "occasionally," and "rarely" related to

"involvement with teacher." Why "active participation" is equated with "involvement with teacher" or how "frequent" was differentiated from "occasionally" is not given. The credibility of the conclusions rest directly on the appropriateness and adequacy of the author's procedure and instruments. In their absence, proceeding to the findings and conclusions of the study represents a large step of faith.

However, when the conclusions are studied, the author seems to believe that a correlation indicates cause-effect relationship. Within the limitations described above, he has some very interesting correlations. He assumes, however, that on one hand level of achievement and level of involvement with teacher are influenced by personality profiles and yet in his conclusion he suggests that we change instruction in order to change a pupil's personality. While this may be a productive strategy, the design of his study does not permit that generalization. In no way did he vary the instructional treatment so that this effect in either achievement, involvement or personality characteristics of pupils could be observed.

Explorative correlational research may indeed be productive in indicating some possible relationships between variables. It certainly does not demonstrate either their existence or the strength of their existence. Based on the outcomes of this reseach report and others that could be cited, we now need studies that will examine the variables in an experimental setting. The findings of these studies properly translated into direct implications for the teaching of science will enhance the practical value of such research for the science teacher.

EJ 121 404

Raven, Ronald J. and Robert Guerin, "Quasi-Simplex Analysis of Piaget's Operative Structures and Stages." Science Education, Vol. 59, No. 2:273-281, 1975.

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

Expanded Abstract and Analysis Prepared Especially for I.S.E. by N. Eldred Bingham, University of Florida.

Purpose

The purpose of this study was to analyze the hierarchial scheme of seven operative structures defined by Jean Piaget. These operative structures are important because the investigations of Piaget have shown that they determine the form and the function of concept acquisition. As the individual acquires a repertoire of more complex operative structures, he is capable of acquiring more complex concepts.

This study assessed the sequential hierarchial interdependence of the following operative structures in the order that they are listed: classification, seriation, logical multiplication, compensation, proportional thinking, probability, and correlational thinking. Piaget does not suggest this order, but the possibility of the order can be inferred from his studies.

The hypotheses of the study were:

H1: There is a hierarchial structure among the concrete operational stages III A and B and the formal operational stages IV A and B.

H2: There is a hierarchical structure among the logical operations of classification, seriation, logical multiplication, compensation, proportional thinking, probability and correlational thinking.

H3: There is a factor describing stage III A operative structures and another factor describing stage III B, IV A and IV B operative structures.

Rationale

The progressive acquisition of complex concepts is related to the acquisition of progressively more complex operative structures. If there is a sequential hierarchial interdependence of the operative structures in the order they are listed: classification, seriation, logical multiplication, compensation, proportional thinking, probability, and correlational thinking; then the K-12 science curricula should consider this sequence of operative structures in designing instructional sequences for science concepts.

The operations used in this study are found in the concrete and formal stages described by Piaget. The concrete stage III contains operations which are carried out on the objects themselves. These operations result from the coordination of the actions of combining, dissociating, ordering, and the setting up of correspondences which then require the form of reversible (compensating) systems. The operations belong to the logic of classes and relations but they do not take into account the totality of possible transformation of classes and relations. Stage III A contains classification and seriation operations.

Stage III B contains logical multiplication and compensation operations. Logical multiplication is the construction of correspondences between at least two sets of variables. Compensation is the balancing of two or more changeable variables. Both of these operations involve the coordination of variables within and between subsystems. The formal stage (IV) contains operations which can be combined into a structured whole. The novel feature of this stage is the ability to reason by hypothesis. Hypothetico-deductive reasoning is characterized by the possibility of using any sort of data as purely hypothetical, and reasoning correctly from them.

Proportional and probability thinking are found in stages IV A. Proportional thinking involves the construction of ratios whereas probability thinking constitutes the construction of a ratio of the frequency of occurrence of events. Correlative thinking is found in Stage IV B. Correlative thinking is the construction of a rule describing the relationship between two sets of events that have a probability attached to their occurrence. In many ways, it can be considered to combine the two operations of Stage IV A. Piaget's formulation of these stages indicates that they are sequentially interdependent. This study sought to determine if the four substages previously described constitute a hierarchy of related components.

Research Design and Procedures

The study involved a total sample of 896 students, age range from 8 to 19 years, comprised of 4 subsamples. Subsample one consisted of 129 children from two core area elementary schools in the Philadelphia School District. A total of 43 children were randomly selected from each of the fourth, fifth, and sixth grade combined classes of the two schools. Subsample two consisted of 220 children from an elementary school in the Cheektowaga School District in New York. The area is a suburb of Buffalo with children from families predominantly in the middle socioeconomic class. Subsample three consisted of 424 students from the Salamanca School District in Salamanca, New York. This is a rural area in western New York. Classes within a grade level were randomly selected for this study. Subsample four consisted of 123 black male and female college freshmen at Clark College in Atlanta, Georgia. Their economic and social backgrounds were diverse. These students were all in a remedial program designed to provide them with knowledge requisite for admission to a four year college program.

The Raven Test of Logical Operations (RTLO) contains a total of 68 items. It has the following number of items per logical operation: classification, sixteen items; seriation, eight items; logical multiplication, twelve items; compensation, eight items; proportional thinking, six items; probability, nine items; and correlation, nine items (1). The range of items for concrete stage III A is 0 to 24. The concrete stage III B has a range of 0 to 20 items. The formal stage IV A has a range of 0 to 15 and stage IV B has a range of 0 to 9 items.

The RTLO is divided into three parts. Each part requires 45 minutes for administration. One part of the RTLO was administered to the same subjects by the same person on each of three consecutive days. All subjects were tested in groups and were assigned scores in the same manner.

Guttman's radex theory provides a quantitative model that is parallel to Piaget's qualitative model (2). This simplex part of the radex theory possesses a simple order of complexity.

The perfect simplex consists only of common parts and the quasi-simplex consists of both common and unique parts. Schoeneman used this model to develop an algorithm to estimate the simplex loadings of a quasi-simplex from a set of correlations without having to arrange the variables in a proper simplex order. The model developed by Schoneman provided the basis for the method used to investigate the first two hypotheses (3, 4).

The third hypothesis was investigated by the use of Alpha Factor Analysis (AFA). AFA is designed to investigate the structure of a universe of variables from a sample of that universe.

Findings

The predicted order of complexity was the same as the order of the stages. The mean values for the concrete III A and III B stages and the formal IV A stage ranged between 50 and 70 percent of the maximum value. However, the mean value for the formal stage IV B was but 18 percent of the predicted maximum value. The very small difference became the coefficient for the concrete stage III B and the formal IV A. This suggested that these variables were of equal complexity. The average of the absolute difference between the estimated correlations was .08. Based on these results, the hypothesis of a hierarchial structure among the concrete IV A, concrete III B, formal IV A, and formal IV B stages was accepted (H1).

The same quasi-simplex procedure was applied to the logical operations variables (classification, seriation, logical multiplication, compensation, proportional thinking, probabilities, and correlation). The rank order of the simplex coefficients associated with the logical operation variables was not the same as the predicted order. Consequently the second hypothesis of a hierarchial structure among the logical operations variable was rejected (H2).

The independent solution derived from the alpha patterns matrix was used to test the third hypothesis that "there is a factor describing stage III A operative structures and another factor describing stage III B, IV A and IV B operative structures" (H3). The first factor had a generalizability coefficient of 0.91; the second factor a coefficient of 0.72. On the basis of these results, the third hypothesis was accepted.

Interpretations

The results of the study show that the two concrete substages, III A and III B, and the two formal substages, IV A and IV B, of cognitive development described by Piaget are hierarchically related. The concrete operational stage, III A, is the least complex stage of those studied. This stage is concerned with additive arrangement type operations. The concrete stage, III B, and IV A, formal stage, have an equivalent amount of complexity. Formal stage IV is the most complex stage. The last three stages are similar because they involve the coordination of two or more variables.

These findings are not in complete accordance with Piaget's theory. According to this theory, each stage is dependent upon the previous stage and the stages are hierarchially sequenced. It would be expected that stage III B would precede stage IV A. Stage IV A can be considered to be more complex than stage III A because the operations within stage IV A assign numbers to the variables. However, stages III B and IV A are about the same point on a complexity scale of cognitive development. One explanation for this is that the four operations develop at about the same time, after the individual has acquired classification and seriation operations. In addition to this point, each of the four operative structures in stages III B and IV A coordinate two changing variables and thus they can be considered to be quite similar in the way that they are used. The concept acquisition hierarchy that was initially suggested in this study is not verified. Compensation and logical multiplication need not precede probability and proportional thinking. However, the findings of the study do show that correlational thinking (stage IV B) is dependent upon the previous three stages for its development and that stage III A operations (classification, seriation) are requisite to the development of higher order operations.

Science instruction need not use Piaget's operative structures solely as a developmental scheme that places concept acquisition in a lockstep framework. The results of this study indicate that many of these operative structures develop simultaneously. Although these operative structures are related within a stage, they are not sequentially dependent in a linear fashion.

The operative structures provide a means for reorganizing a science concept in many different ways. The capability of restructuring a concept permits it to be applied to situations that vary in the quality of their patterns or organization. More important, the ability to restructure a concept and apply it to different situations enhances the pupil's ability to generalize and retain the

concept. The operative structures domain investigated in this study can be thought of as enhancing the development of science concepts and extending the use of these concepts to a variety of different situations.

ABSTRACTOR'S ANALYSIS

Raven's study extends Piagetian research to a large sample of fourth, fifth and sixth grade children in four widely separated geographic locations, each representing a different socioeconomic class. It also applies statistical measures to test the probability of hypotheses being sustained under widely varying conditions. These widely varying conditions as well as use of random techniques make possible the extension of the results to other situations. Furthermore, in his previous research Raven had developed a valid and reliable test, entitled Raven Test of Logical Operations, (RTLO). This instrument administered in three 45 minute sessions, gave the required data for the analysis required. His choice of statistical method is in my view an appropriate one to use when one wishes to test the requisite seriation of logical operation as hypothesized.

Realizing that logical operations of classification, seriation and logical multiplication are concerned with the arrangement of events and objects by a property and that compensation, proportional thinking probability and correlational thinking are concerned with the coordination of changing variables it appears that the use of the Alpha Factor Analysis of Schoneman was appropriate for testing the third hypothesis.

It seems clear to me that this research, along with others, shows the usefulness of Piaget's theory as it applies to children up through Piaget's concrete III A, but that beyond this stage, as the authors suggest, and as Novak states in his editorial comment, "A more diffuse growth and use of concepts may better characterize cognitive development. The results are consistent with Ausubel's theory of learning in that the level of cognitive functions are importantly influenced by the extent of specific, relevant concepts available in the learner's cognitive structure."

REFERENCES

1. Raven, Ronald, "The Development of a Test of Piaget's Logical Operations," Science Education, Vol. 57, No. 3:377-385, 1973.

2. Guttman, Louis, "A New Approach to Factor Analysis: the Radex," in Lazersfeld, P. (Ed.). Mathematical Thinking in the Social Sciences. Chicago: The Free Press, 1954.

3. Schoneman, Peter, "Fitting a Simplex Symmetrically," Psychometrica, Vol. 35, No. 1:1-21, 1970.

4. Guerin, Robert. Simplex. Buffalo: Faculty of Educational Studies, State University of New York at Buffalo, 1974.

EJ 023 201

Scott, Norval, "Strategy of Inquiry and Styles of Categorization: A Three-Year Exploratory Study." Journal of Research in Science Teaching, Vol. 7, No. 2:95-102, 1970.
Descriptors--*Cognitive Processes, Classification, *Cognitive Tests, *Elementary School Science, *Inquiry Training, *Learning, Verbal Ability

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Leslie R. Allen, University of Hawaii.

Purpose

The stated purpose of this study was to determine answers to the following questions:

1. Would an Inquiry program have a continued effect on children's behavior after the novelty of the situation had passed?

2. Would the verbal behavioral changes in Inquiry children in a three-year study be traceable to the elements of the strategy emphasized during this program?

Rationale

A 1966 study by Scott found that the verbal behavior of children involved in the Detroit Inquiry program differed substantially from that of conventionally taught pupils. Scott's study lasted only one year. The present investigation was concerned with determining the long-term effects of such an Inquiry program.

Research Design and Procedure

In September, 1962, the Detroit Inquiry Program was introduced to a group of 25 fifth graders in an elementary school in Detroit. The same approach was continued with the same children until June, 1965. This experimental group received 100 minutes of Inquiry training per week in the science class, and they were tested with the Sigel Cognitive Styles Task in September, 1962, June, 1964, and June, 1965.

There were three control groups, one for each of the above testing periods. These control groups, consisting of children who had not been subjected to Inquiry training, were matched with the experimental group using the criteria of sex, socioeconomic level, and achievement or mental maturity. Two of the three control groups (Grades 5 and 6) were from the same school. The third (Grade 7) was from a junior high school group of similar racial and ethnic background, and with similar achievement rating. Each control group consisted of 25 pupils. Experimental and control groups were tested at approximately the same time.

In the Detroit Inquiry Program the child is faced with a series of events which seem to be at odds with his perception of the environment. His task is to obtain enough information so that he can explain why things happened as they did. He does this by asking the teacher questions, answerable by "Yes" or "No," about the experiment. This leads to an attempt at an explanation by the child. If the explanation is correct, the teacher tells him so. The teacher then checks understanding by posing several critical questions. Between problem-solving sessions, strategy sessions are held in which the teacher and the pupils analyze the questions recorded from the problem session in order to increase the efficiency of the Inquiry procedure.

The test instrument, the Sigel Cognitive Styles Task (SCST), consists of a booklet of 35 cards, each composed of three pictures of familiar objects. The child selects pairs of pictures from each card and writes as many reasons as he wants for his choices. His written labels are then coded into the following groups:

DPW visible detail of objects; for example, B and C have helmets

DW global aspects of stimuli; for example B and C are both men

R thematic relationship; for example, fireman drives a truck

C1 functional statements; for example, B and C are serving the city

C2 taxonomic labels; for example B and C are mammals

C3 inferred attributes; for example, B and C are tired.

Two other measurements were obtained, the stimulus shift (SS), how often the child shifted from one pair of pictures to another within each card, and the category shift (CS), the number of times he shifted from one category to another in labeling the first combination of pictures for each card.

Findings and Interpretations

Chi-square was used to test for significant differences between the experimental and the control groups for each of the 8 SCST categories at each of the three testing periods, that is, a total of 24 null hypotheses. Significant differences in favor of the experimental group were noted in both 1964 and 1965 for the following 4 SCST categories: DPW, C3, SS, and CS.

In other words, with respect to the above four categories, there was no sign of regression among Inquiry pupils over the three year period of the study. The Inquiry program also seems to have had a persistent effect on styles of categorization.

Bar graphs presented show differences in the kinds of DPW labels used by Inquiry and control children during the 1965 testing. Inquiry children were shown to be more inclined to label small visible details, both human and inanimate, and to describe the background configurations of the illustrations than were the comparison children; on the other hand, the comparison children favored labeling items on the basis of material, posture, position, covering, and size or shape.

Analysis of the C3 scores shows that the Inquiry pupils used a larger proportion of labels referring to inherent physiological, physical, and morphological characteristics, whereas the control pupils responded with a greater percentage of emotional and geographical/locational labels. It appears, then, that the pupil who has learned to focus on the non-manifest, factual details of a science problem will be less likely to make inferences as to happiness, sadness, or other emotional qualities, and more likely to consider inherent physical and physiological characteristics, such as blood, muscle, or motors, than will those pupils who have not had this type of experience.

These trends find an explanation in the fact that the Inquiry child must penetrate to the invisible, inherent characteristics of a problem demonstration, detecting the forces and energies involved, and space/time continua, before he is likely to have success in problem-solving. Another parallel appears, then, to exist between the Inquiry strategy and styles of categorization: the child who inquires about the invisible, inherent, aspects of a problem may prefer to interact with the invisible characteristics of the environment.

Finally, longitudinal trends were examined <u>within the Inquiry responses alone</u>. The percentage of DPW responses to inanimate parts and background configurations increased from 1962 to 1965, and there was a decrease in the amount of attention given to kind of material, covering, and shape. C3 labels showed an increasing percentage of responses again to physical, physiological, and morphological characteristics, and an increased willingness to make value judgments. On the other hand, there was a decrease in labels of an emotional and geographical nature. This information reinforces the parallel between DPW and C3 labelling modes discussed earlier. In other words, compared to <u>itself</u> before the Inquiry strategy experience and compared to children matched on criteria other than Inquiry, the Inquiry group has indicated differences after Inquiry exposure that are not obviously explainable except in the light of changes that might be effected by the program itself.

Summarizing then, the Inquiry strategy appears to have had a continuing effect on the verbal behavior of this group of children over the three-year testing period. The children exposed to the technique changed in several measurable ways: verbal fluency and flexibility were increased, attention to detail became more acute, inferences as to invisible attributes showed a strong trend away from the emotional and locational responses, and towards the inherent classificatory attributes. Each of these changes can reasonably be traced to a specific emphasis of the Inquiry strategy used in this program.

ABSTRACTOR'S ANALYSIS

The main criticisms of this study lie in the experimental design and the statistical treatment. A number of questions need to be asked: What was the size of the population from which the experimental and control groups were selected? Were the 25 experimental children in one class? If not, how were they chosen? Why such a low N? No mention is made of the attrition rate of the experimental group over the three-year period of the study. Were the drop-outs replaced? If so, how? No explanation is given for the use of a new control group each year. Had the same group been used, a trend analysis of results over three years would have been possible.

No attempt has been made to randomize either selection of pupils or assignment of pupils to treatments. If matching is desired, it is preferable to match pupils on one or more criteria and then randomly assign them to treatments.

The control groups were described as "children who had not experienced the Inquiry process." But what were these children doing during their science classes? More information is needed here to determine the degree to which the experimental treatment differed from that given to the controls. Furthermore, what amount of time per week did the controls spend on science?

Although the test instrument used was adequately described, no reliability or validity coefficients are given.

The abstractor wonders why analysis of variance (or covariance), or the Fisher Probability Table, was not used for data analysis rather than the Chi-square test. No explanation is offered.

Since the same test was used as a post-test for three subsequent years, a legitimate question is: How much of the gain in achievement was due to taking the pre-test and how much was due to participation in the Inquiry program?

This paper is a useful addition to Inquiry Literature. Unfortunately, replication is made difficult, if not impossible, by virtue of the author's failure to provide an adequate description of the pupils involved. It would be necessary to know age, sex, socioeconomic level, and achievement or mental maturity. The results from upper socioeconomic level, average to high I.Q. pupils (or low socioeconomic level, low I.Q.) may differ markedly from those that would be obtained from a different pupil population.

INSTRUCTION

EJ 023 083
EJ 023 084

Citron, Irvin M. and Cyrus W. Barnes, "The Search for More Effective Methods of Teaching High School Biology to Slow Learners Through Interaction Analysis, Part I. The Effects of Varying Teaching Patterns. Part II. The Effects of Various Constant Teaching Patterns." Journal of Research in Science Teaching, Vol. 7, No. 1:9-28, 1970.

Descriptors--Academic Achievement, *Biology, Classroom Observation Techniques, *Instruction, *Interaction Process Analysis, *Slow Learners, *Secondary School Science, Teaching Styles

Expanded Abstract and Analysis Prepared Especially for I.S.E. by James Reed Campbell, St. John's University.

Purpose

The purpose of this study was to determine which interactive strategies are most productive (achievement) for slow learners in biology.

Hypotheses:

1. A high I/D ratio early in a course of biology for slow learners followed by a lower I/D ratio later in the course increases achievement.

2. A low I/D ratio early in a course of biology for slow learners followed by a higher I/D ratio later in the course increases achievement.

3. A constant medium I/D ratio throughout a course of biology for slow learners leads to higher achievement in concept formation than does a change in I/D ratio in either direction from the start to the end of the course.

4. A positive relationship exists between the extent of the use of indirect verbal interaction in biology classes for slow learners and their achievement in understanding biology concepts.

5. A positive relationship exists between the extent of the use of indirect verbal interaction in biology classes for slow learners and their achievement in increased ability to solve biology problems.

6. Insofar as the measurement of total achievement is based upon a substantial percentage of problem-solving test items, there is a positive relationship between the extent of the use of indirect verbal interaction in biology classes for slow learners and their total achievement in biology.

Rationale

This research involves the utilization of the empirically based indirect/direct methodology developed by Flanders, et. al. The research also concerns evaluating methodologies involved with slow learners and programs specifically developed for this group of students.

Research Design and Procedures

This study is experimental in nature because it utilizes six distinct interactive treatments which are applied to groups of tenth-grade slow learner students in biology. The dependent variables involve achievement in terms of concept formation and the ability to solve problems in biology.

Research Design:

Treatments (First Phase-Fall)	Treatments (Second Phase-Spring)
X_1 Varying Interaction O_1 O_2 O_3 High I/D to Low I/D N = 3 classes	X_4 Constant Interaction O_4 O_5 O_6 High I/D N = 3 classes
X_2 Varying Interaction O_1 O_2 O_3 Low I/D to High I/D N = 3 classes	X_5 Constant Interaction O_4 O_5 O_6 Low I/D N = 3 classes
X_3 Constant Interaction O_1 O_2 O_3 Intermediate I/D N = 3 classes	X_6 Constant Interaction O_4 O_5 O_6 Intermediate I/D N = 3 classes
O_1 O_2 O_3 First semester tests	O_4 O_5 O_6 Second semester tests

O_1 O_4 Total Achievement

O_2 O_5 Items relating to biological concepts (Concept Formation)

O_3 O_6 Items relating to solving biological problems (Problem Solving)

Class Size $\bar{X}$ = 26 students

Measurement Instruments: The ten-category Flanders Interaction Analysis System (FIAC) was used. This system quantifies the affective climate of classes. The curriculum used was the BSCS slow learner biology program, Patterns and Processes. The tests used in the study were the tests provided by the BSCS program.

Duration of Study: The fall semester was used for the first three treatments, and the following spring semester was used for the next three treatments.

Procedure

Teachers were trained and monitored periodically to make sure they followed the prescribed treatments. After completing the treatments in the first semester, the lowest achieving groups received the constant indirect treatment (X_4), while the highest achieving groups received the constant direct treatment (X_5). This procedure was used to counterbalance the effects of the first set of treatments. The assumption was made that high indirect ratios would subsequently increase achievement, while low ratios would decrease achievement.

Statistical Methods: The primary method of analysis of the data was to use the differences of the class mean from the group mean. Correlations, chi square tests, and analysis of variance were used to determine the significance of the differences among groups.

Findings

1. A high I/D ratio early in a course followed by a low I/D ratio later in the course resulted in higher achievement in solving biological problems but not in concept formation.

2. A low I/D ratio early in a course followed by a high I/D ratio later in a course did not increase achievement.

3. A constant intermediate I/D ratio throughout a course led to higher achievement in concept formation than did a change in I/D ratio in either direction.

4. Constant indirect methodology (high I/D ratio) throughout a semester did not produce a positive relationship with achievement in concept formation.

5. Constant indirect methodology did produce significantly greater achievement in problem solving and in total achievement (60 percent of test questions were classified as problem solving).

Interpretations

Consistency in indirectness seems to be positively related to higher achievement in solving problems in biology but not necessarily in specific content achievement. Varying ratios within specified high and low levels does not result in higher levels of achievement for slow learners.

ABSTRACTOR'S ANALYSIS

This study is one of the very few interaction studies which is experimental in nature. A specified long-term treatment was applied to groups of slow learners and some counterbalancing of treatments was used. The intermediate interactive treatment could be classified

as a control group if the I/D ratios were in the neutral range. Unfortunately, the study is not a true experiment because the selection of the sample was not randomly assigned to the treatments. To alleviate the problem, the researchers could have utilized a pre-post test design with analysis of covariance to adjust for any pre-test inequalities. Pre-test data are necessary when the researcher is forced to use intact groups because it must be established that the groups are equivalent at the beginning of the experiment. Without such information we cannot be certain of the effect of the treatment on the post-test results.

Another problem in not randomly assigning the students is the possibility of some groups containing youngsters with higher IQ's. The IQ range of the classes was 70-100, and it is possible that some classes might have had a preponderance of 90 IQ youngsters, while another group might have had large numbers of 80 IQ students. Such differences could easily contaminate the findings if they are not monitored closely.

Another design problem concerns the failure to include a true control group which received the normal instruction provided in these districts. Perhaps all the experimental groups would have scored higher than the control group.

Two questions related to the technical aspects of the research are unanswered. One involves the rationale for using the cumbersome I/D ratio when most researchers have come to use the more precise TRR ratio [Borsch (1)]. The TRR ratio is a modified form of the i/d ratio and has been found to more accurately describe the interactive climate of classes. It eliminates tallies in the teacher's questioning and lecturing categories and is more sensitive to the most critical affective elements of the FIAC system. Another technical question involves the inter and intra reliabilities of the individual(s) responsible for the coding of the dialogue. Such data are important to include in any research report because coding is extended over a period of time, and it must be demonstrated that the coding is performed within accurate and consistent limits.

The design of the experiment was really separated into three groups with three separate treatments given to each group. In a sense, the research represents two replications of one experiment each semester. In all the analyses the groups are always kept separate and the treatment across groups is never summarized in diagram or table data. Such data would have made the findings more understandable.

The one major difficulty with the large volume of research done with the Flanders instrument is the lack of a developing theoretical framework to fit the individual studies. This study represents another empirically based investigation which adds much useful information but does not lead to the development of the needed theoretical constructs. What is needed is a model or theory which begins to tie together the many individual research efforts in this area. This study, with its experimental counterbalancing design and its successful utilization of long-term interactive methodologies, is a step

forward. What is needed are a second generation of experimental interactive studies constructed from within a theoretical framework.

REFERENCES

1. Bosch, Albert C., "Relationship of Teaching Patterns to Indices of Classroom Verbal Interaction Behavior: A Further Analysis and Synthesis of Classroom Verbal Interaction Data Using Descriptive Indices of Behavior and Teaching Patterns Analysis," University Microfilms, Ann Arbor, Michigan, 1972.

EJ 057 451

James, Robert K., "A Comparison of Group and Individualized Instructional Techniques in Seventh Grade Science." Journal of Research in Science Teaching, Vol. 9, No. 1:91-96, 1972.

Descriptors--Achievement, Critical Thinking, *Group Instruction, Instruction, *Individualized Instruction, *Programmed Instruction, *Secondary School Science, Scientific Enterprise

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

Purpose

Individualizing instruction in science teaching has long been discussed but found to be elusive in practice. The Intermediate Science Curriculum Study (ISCS) has attempted to provide a practical and workable strategy for junior high school science teachers to individualize their science instruction. This study was designed to compare student outcomes between classes which experienced individualized and group instruction.

Rationale

The Intermediate Science Curriculum Study has developed a set of materials to be used for individualized instruction. The author felt that certain basic questions about the use of those materials had not been answered.

1. What effect will their use have on the interest of the students in science? "Studies such as one by Woodruff (8) indicate that with programmed materials used over a year-long course, considerable reduction in interest in science results.

2. Will the student be able and willing to assume the responsibility necessary for him to achieve at satisfactory levels?

3. Can students, who are willing to accept responsibility for their learning, function well in this relatively unstructured classroom environment? Reports of visitors to ISCS classrooms have suggested that they cannot.

4. How will different ability levels react to this treatment? Will the lack of typical classroom structure and dominance by the teacher impair the achievement and interest of certain students?

5. Can a teacher manage this instructional technique without detriment to the achievement of his students?

In an effort to find answers to these questions, this study was designed to compare student outcomes between classes which experienced individualized and group instruction. ISCS materials were not available for conversion to group techniques at the time this research began, so it was decided that an established course outline would be rewritten so that it would employ an approach as similar as possible to that used in ISCS. Materials originally developed for the seventh grade course, "Matter," at University Schools, University of Iowa, were selected because the philosophy and goals were similar to those of ISCS and because of the researcher's familiarity with the "Matter" course.

Research Design and Procedure

The sample consisted of seventh grade science classes at University Schools, University of Iowa. Both treatments were the same with respect to the content outline, sequence, student experiments, and content and process objectives. However, the individualized treatment allowed faster students to go beyond the core sequence, while slower students were not required to complete it. The 60 students were heterogeneously grouped into the two treatments by randomly assigning half of the upper, middle, and lower thirds of the entire group to each treatment. This was accomplished by using the total scores on the general aptitude scale, the Academic Promise Tests (1). The Seventh Grade Matter Final (SGMF) was designed to measure facts and concepts which were common to both treatments. The Read General Science Test was used to measure achievement of facts and concepts in broad areas of science, i.e., beyond the scope of the content that was common to both treatments. All pre-tests were administered in September, and the post-tests were administered in May.

The statistical technique was an analysis of covariance within a treatments by levels design. Both techniques increase precision over a simple randomized design, and have been used simultaneously in this study. The use of the treatments by levels technique had two basic advantages in this study. It enabled the testing of hypotheses about the achievements of different aptitude levels within the two classes, and it increased the likelihood of meeting two of the basic assumptions of the test of significance for analysis of covariance; a normal distribution and equality of variance (4). Pre-test scores were used as covariates in each instrument except the Seventh Grade Matter Final. Since this test was not available for use as a pre-test, the composite score of these students on the Iowa Tests of Basic Skills (ITBS) was used.

Null hypotheses were stated regarding the difference between the means of the two treatment populations with respect to:

1. the achievement of facts and concepts as measured by the Seventh Grade Matter Final (SGMF), and Read General Science Test (RGS) (6);

2. the development of critical thinking skills as measured by the total test score on the Watson-Glaser Critical Thinking Appraisal (WGCTA) (7);

3. their understanding of the nature and methods of science as measured on the subtest and total test scores by the Test on Understanding Science (TOUS) (2);

4. their subject preference as measured by the Prouse Subject Preference Test (PSPT) (5).

Findings

All four null hypotheses must be retained at the 0.05 level. Additionally, analysis of the TOUS subtests indicates that null hypotheses for subtests I and II could be retained. Subtest III, measuring the methods and aims of science, could be retained.

Interpretations

The author proposes the following interpretations. Failure to be able to reject the hypotheses must cause one to be cautious in generalizing these results. However, the following tend to be supported by this research:

1. The students' interest in science did not tend to be different between the two groups. This does not agree with the results reported for year-long courses employing programmed materials.

2. Failure to find differences in the achievement between the two treatments tends to support the idea that students in the individualized treatment are able to assume responsibility for their learning, and profit from an environment which has been judged by observers as "chaos."

3. As to interaction of the individualized treatment with ability levels, this study did not provide evidence to support the contention that poorer students are more apt to profit from individualized instruction.

4. The teacher-investigator did find that this classroom can be managed.

ABSTRACTOR'S NOTES

The need for studying various types of methods is not questioned. There is a critical problem of studying methods over a short period of time. It seems important to extend studies of this type over a longer period of time before having any confidence in generalizing the results.

One must question the advisability of the teacher-investigator teaching both groups with differing methods and being able to sharply differentiate the methods in the two groups. It is assumed that the ISCS materials were taught as planned by the developers. There is no description of the instructional methods used for the "group" instruction. Were there demonstrations or any laboratory exercises for the students? From the results of the testing it is tempting to say that the research might indicate that the teacher is the dominant factor in this case rather than the materials or the methodology of the program. The teacher-investigator developed the "Matter" course with the philosophy and content of ISCS. The philosophy and content evidently came through consistently.

The author stated in the introduction that there were five questions to be investigated in this research. There is little similarity between the stated questions and the null hypotheses. Indeed, null hypotheses 2 and 3 are not found within the original questions.

One of the incidental comments made by the author is interesting to note. "Both treatments were asked to evaluate how easily they were distracted from their work during the class. Two-thirds of the class having group instruction described themselves as 'easily distracted,' while about half of the individualized treatment group reported this. This is somewhat the opposite result that had been predicted by many experienced teachers." Are there any data to support these statements? It seems likely that students who must assume more responsibility for their own work (individualized) would be less distracted unless the teacher would not relinquish the learning responsibility to the students.

Research regarding methodology is important and needs to be investigated. In order to gain some insights, the questions need to be clearly stated so that precise procedures can be developed.

REFERENCES

1. Bennet, George K. and others. Academic Promise Tests. New York: The Psychological Corporation, 1965.

2. Cooley, William W. and Leo E. Klopfer. Test on Understanding Science Manual. Arlington, Virginia: ERIC Document Reproduction Service, P.O. Box 190, 1961. ED 010 091

3. Heath, Robert W., "Pitfalls in the Evaluation of New Curricula," Science Education, Vol. 64:211, 1962.

4. Lindquist, E. F.. Design and Analysis of Experiments in Psychology and Education. Boston: Houghton Mifflin, 1953.

5. Prouse, Howard L., The Construction and Use of a Test for Measurement of Certain Aspects of Creativity in Seventh Grade Mathematics, Dissertation, University of Iowa, 1964.

6. Read, John C. Read General Science Test Manual. Chicago: Harcourt, Brace and World, 1965.

7. Watson, Goodwin and Edward Glaser. Watson-Glaser Critical Thinking Appraisal Manual. New York: Harcourt, Brace and World, 1964.

8. Woodruff, A. Bond. Methods of Programmed Instruction Re Student Characteristics. Arlington, Virginia: ERIC Document Reproduction Service, P. O. Box 190, 1965. ED 003 335

EJ 065 059

Jones, Paul L. and Jacob W. Blankenship, "The Relationship of Pupil Control Ideology and Innovative Classroom Practices." Journal of Research in Science Teaching, Vol. 9, No. 3:281-285, 1972.
Descriptors--Biology, Classroom Research, *Discipline, Science Course Improvement Project, Secondary School Science, *Teacher Characteristics, *Teaching Styles

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

Purpose

This study was designed to investigate whether or not biology teachers who have a humanistic orientation toward pupil control exhibit practices recommended by BSCS to a greater extent than do biology teachers who have a custodial orientation toward pupil control. The specific hypothesis tested was as follows: "There is no significant difference in the classroom teaching practices recommended by the BSCS between teachers who have a more humanistic pupil control ideology and those teachers who have a more custodial pupil control ideology."

Rationale

Support for the investigation centered around previous findings reported by Willower (4) and Willower and Jones (6) and a belief that innovative teaching techniques such as those suggested by the BSCS staff require a classroom environment in which students are free to become active participants in the teaching-learning process. The research findings pointed out that teachers' concern for pupil control and fear of being categorized as soft disciplinarians have restricted the implementation of innovative teaching procedures which are consistent with the newer science curricula.

The major assumption underlying the study was the acceptance of the validity and reliability of the Pupil Control Ideology Form (5) and the Biology Classroom Activity Checklist (3).

Research Design and Procedure

The Pupil Control Ideology Form (PCI) was administered to 68 biology teachers who were selected by means of a stratified, randomized process. The instrument, consisting of 20 items scored on a Likert-type scale, provided an operational measure of the concepts of humanism and custodialism. It measures teacher ideology concerning these concepts and not the actual teacher classroom behavior. The range of possible scores was 20 to 100, with low scores representing a more humanistic orientation toward pupil control and high scores representing a more custodial orientation toward pupil control.

Teacher classroom practices were measured by administering the Biology Classroom Activity Checklist (BCAC) to 2040 students who were enrolled in the first biology classes of the day for the 68 biology teachers. The BCAC provides a description of seven classroom practices from a student's point of view. These include role of teacher, student participation, use of curriculum materials, tests, pre-laboratory practice, laboratory practice and post-laboratory practice. The instrument is composed of 53 items; 26 are positively and 27 are negatively related to teacher practices recommended by BSCS.

The 25 higher scores and 25 lower scores on the PCI were employed in selecting two comparison groups of teachers. The group with the higher scores was designated as being custodial in their pupil control ideology, while the group with the lower scores was designated as being humanistic in their pupil control ideology. A comparison between the higher and lower scoring groups was accomplished by applying a single classification analysis of variance to mean student scores for each class on the BCAC. Fifteen hundred of the 2040 student responses on the BCAC were used in the analysis.

Findings

The findings reported by the investigators were as follows:

1. a significant difference at the 0.05 level was found between the humanistic and custodial groups of teachers for total scores on the BCAC;

2. a significant difference at the 0.05 level was found between the humanistic and custodial groups of teachers for role of teacher, use of curriculum materials, texts, and post-laboratory practice; and

3. no significant difference at 0.05 level was found between the two groups for student participation, pre-laboratory practice, and laboratory practice.

Interpretations

The investigators concluded that teachers who had a more humanistic pupil control ideology would be preferred to work with the BSCS program, because they exhibited classroom practices recommended by BSCS to a greater extent than their counterparts who had a more custodial pupil control ideology. Specific examples of the classroom practices exhibited more often by teachers with a more humanistic orientation toward pupil control were as follows:

1. asks questions which cause students to explain meaning of textbook material, think about previous learning, and examine evidence behind textbook material;

2. encourages students to question the textbook and substantiate questions by reading additional sources;

3. utilizes open-ended test questions; and

4. discusses and analyzes results obtained in the laboratory in post-laboratory situations.

The investigators also concluded that a relationship did not appear to exist between classroom practices recommended by BSCS and pupil control ideology when the teacher was not a central figure in teacher-student interactions.

ABSTRACTOR'S ANALYSIS

For the most part, the "newer" science curricula of the 1960's have not been implemented into the elementary and secondary schools in the manner intended by the developers of the curricula. Reasons for this phenomenon are many but no doubt include the lack of understanding of the relationship that may exist between teacher classroom practices recommended for proper curricula implementation and various psychological factors, including teacher orientation toward classroom control. The study by Jones and Blankenship represents an investigation of an area that has been, and is, in sore need of attention. In fact, the major contribution of the investigation may be that it focused attention on one of the important psychological factors which is likely to affect classroom practices. If curricula are ever to be implemented in a manner consistent with the philosophy of the developers, it will be necessary to gain further insight into teacher control ideology and how this ideology manifests itself in terms of teacher classroom practice.

One of the problems associated with reporting an investigation is how to stay within the length acceptable by the publisher and still include enough information to make the report as meaningful as possible for the readers. In many instances such space limitations cause the investigators to present only a sketchy view of the instruments used along with an appropriate reference, since they were likely to have been developed by other investigators. This situation appears to be the practice adopted by Jones and Blankenship. Such a practice is acceptable provided the instruments are easily accessible and/or well known by the potential readers. However, the PCI has not been widely used in science education research; it is not easily accessible; and, as far as the reviewer could determine, has not been described or evaluated in one of the standard reference books on mental measurement. The research report would have been more meaningful if the investigators had included a more detailed description of the PCI in their research report.

The investigators should also have provided additional information on the PCI scores obtained by the teachers. They stated that the upper and lower scores were selected and designated as more custodial and more humanistic. The practice of selecting extreme scores is not a questionable practice, but without the scores, the reader cannot

interpret with any accuracy the actual nature of the two groups. It is entirely possible to select upper and lower groups both of which are really custodial or humanistic. Examples of this problem have been observed in other studies involving interaction analysis where extreme groups were actually direct rather than being indirect and direct in their classroom behavior (1). It also would have been useful to know whether or not the group scores were significantly different. In other words, without further data the reviewer cannot be certain that more custodial and more humanistic groups of teachers were being compared.

An additonal report problem centered around a lack of description concerning the teachers used in the investigation. They were supposedly a stratified random sample, but a sample of what population? How was the sampling accomplished? Were they secondary school biology teachers? Were they all drawn from one district, one state? Without this information, generalizability becomes a problem for the reader.

Any investigation is limited by the validity and reliability of the instruments used for gathering data. This investigation was no exception. In fact, the greatest weakness of the investigation may well have been the acceptance of the assumption of the PCI as a valid instrument for measuring teacher orientation toward pupil control. In the development of the PCI, validity was determined by comparing PCI mean scores of two groups of teachers who had been judged by their principals as being humanistic or custodial. The principals had been asked to rate teachers with respect to teacher ideology concerning pupil control and not on the basis of actual classroom practice (5). Using principal ratings of teacher ideology to validate an instrument appears to be questionble when one considers the inconsistent and conflicting results obtained by other educational researchers who have used principal ratings as criterion measures (1). Further support of this point was made by Remmers in a previous review of research utilizing ratings in education research. He concluded that "the human rater. . . is imperfectly reliable and not often highly valid in his recorded judgments" (2:372). An additional check on validity was made by comparing the PCI scores of teachers from schools whose reputations were humanistic with the scores of teachers in other schools. The authors of the PCI stated with respect to the additional check on validity that "while no statistical analysis was made in this instance, a trend in the expected direction is clearly visible" (5:13). Such procedures can hardly be considered as rigorous checks on validity.

Utilization of a student checklist to obtain measures of classroom practices reveals another area in need of further investigation. Kochendorfer (3) validated the items on his checklist by subjecting them to a panel of judges. This is a common procedure for validating a checklist. It is possible, however, for the items on a checklist to be valid in the sense that they are or are not in agreement with practices recommended by BSCS. Yet, the use of the instrument may not provide a valid description of practices that actually occurred. In this situation, valid measures of classroom practices depended on students' perception of what occurred in the classroom. The reviewer's experience with student checklists does not lend much

credence to the utilization of student checklists for obtaining accurate descriptions of classroom practice. Too many factors may influence students' judgments. For example, ratings may be biased by students' liking or disliking the teacher. Or, if the teacher had conducted a post-laboratory session the day before the checklist was administered, the students might check this activity as being a regular practice when in fact it may have only occurred the one time. Of course, such situations are only speculative, but their possibility points out that more research is needed to determine the relationship that may or may not exist between student scores on checklists, such as BCAC, and systematic observations of actual classroom practices.

In summary, the investigation centered on an area of study which is in need of attention. An understanding of the relationship between teacher ideology concerning student control and classroom practices may be crucial if future curricula are to be implemented in the manner suggested by curricula developers. Analysis of the investigation revealed that researchers should be extremely cautious in their acceptance of the validity and reliability of criterion instruments, especially checklists and instruments validated by principal ratings. The desirability of including more detailed descriptions of research population, instruments and data in research reports was identified. Further analysis revealed the need for the development of more valid instruments for measuring teacher control ideology. The need for further research was also identified concerning the relationship of student perception of classroom practices and systematic observation of classroom practices.

REFERENCES

1. Balzer, A. LeVon, Thomas P. Evans, and Patricia E. Blosser. A Review of Research on Teacher Behavior. Columbus, Ohio: Association for the Education of Teachers in Science and ERIC Science, Mathematics and Environmental Education Clearinghouse, 1973.

2. Gage, Nathaniel L. (ed.) Handbook of Research on Teaching. Chicago: Rand McNally and Co., 1965.

3. Kochendorfer, Leonard. "A Comparative Study of the Classroom Practices and Teaching Rationale of High School Biology Teachers Using Different Curriculum Materials." Unpublished doctoral dissertation, The University of Texas, 1966.

4. Willower, Donald J. "Barriers to Change in Educational Organizations," Theory Into Practice, 2:262, 1963.

5. Willower, Donald J., Terry L. Eidell and Wayne K. Hoy. The School and Pupil Control Ideology. Monograph No. 24. University Park, Pa.: Pennsylvania State University, 1967.

6. Willower, Donald J. and Ronald G. Jones. "Where Pupil Control Becomes an Institutional Theme," Phi Delta Kappan, 45:107, 1973.

EJ 034 670

Okey, James R. and Robert M. Gagné, "Revision of a Science Topic Using Evidence of Performance on Subordinate Skills." Journal of Research in Science Teaching, Vol. 7, No. 4:321-325, 1970.
Descriptors--*Achievement, *Chemistry, *Curriculum Development, *Instruction, Learning, *Secondary School Science, Student Behavior

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

Purpose

The stated purpose of this study was "to investigate the prediction that a way to improve performance on a final task is to provide additional instruction leading to attainment of subordinate skills failed by a substantial number of students."

Rationale

A cumulative learning model has been proposed by Gagné (3) in which learning tasks are structured hierarchially in order that desired learning outcomes will be achieved with a minimum of instruction. The desired learning outcomes are stated behaviorally, describing what it is that the learner is to be able to do. A hierarchy of component tasks is then constructed by beginning with the desired learning outcome and asking, in effect, "To perform this task, what prerequisite or component tasks must the learner be able to perform?" (2) For each task so identified, the same question is asked, thus generating a hierarchy of learning outcomes based on observable prerequisites.

The analysis can begin at any level and specifies what is assumed to come earlier in the sequence of learning tasks. In practice, the component task analysis stops when the curriculum designer believes that it is safe to assume that a given task can be demonstrated by the intended student population.

In Gagné's model, "the primary purpose of the subtask analysis is to guide initial curriculum design to maximize the success for first-draft instructional materials." The work of Gagné and Paradise (3) suggests "three possible explanations when a criterion task has not been mastered: (a) subordinate learning set has been omitted, (b) subordinate learning set has not been retained due to inadequate practice, and (c) insufficient guidance in the subskills has been provided."

Gagné points out that the hierarchy resulting from a task analysis is by no means the single pathway to the criterion task. Instead, it is a description of the hypothesized optimal pathway to be taken by a typical group of students--the most efficient pathway for most students.

Research Design and Procedure

A published learning activity program on solubility product calculations was selected as the curriculum to be revised using the subordinate task analysis described previously. The analysis culminated in the development of a branching four-level hierarchy of fifteen subordinate skills which are inferred to correspond to the simple association-concept-principle-problem solving hierarchy in Gagné's Conditions of Learning (1).

In a separate investigation, a group of students who had studied the published program responded to pairs of test items designed to measure achievement of the criterion task or subordinate skills. If 80 percent of the students either "passed or failed" both of the items in a pair, the items were judged to be equivalent and were used in the construction of four tests: Criterion Pre- and Post-Tests and Subordinate Skill Pre- and Post-Tests.

The subjects were 135 tenth, eleventh, and twelfth grade students from five chemistry classes at a high school in Lafayette, California. "Names of the students in each class were placed in alphabetical order and alternate names were assigned to each of two treatment groups. Attrition from absences and unavailability of I.Q. and mathematics percentile scores" resulted in treatment groups of 49 and 57 students.

Approximately seven 50-minute class periods were required for a treatment group to take the tests and complete the learning program. The Criterion and Subordinate Skills Pre-Tests were given on the first two days, followed by three days on the program, then two days for the Criterion and Subordinate Skills Post-Tests. While the 49-student control group took the tests and completed the published program, the 57-student experimental group did a series of laboratory investigations on titration and acid-base indicators. Items on the Subordinate Skills Post-Test were analyzed to identify "skills which had been failed by many students." Twenty additional frames, consisting of instruction in 10 of the 15 subordinate skills, were written and added to the original 50-frame program.

Findings

Mean scores on the 7-item Criterion Post-Test increased from 3.02 for students using the published program to 3.84 for students using the revised program. A one-way analysis of covariance showed that differences in the criterion test score gains favored the experimental group ($F = 19.45$, $df = 1/100$, $p < .001$) using pre-treatment measures of I.Q., mathematics percentile, Criterion Pre-Test and Subordinate Skills Pre-Test as covariates.

Comparison of the two treatment groups indicated that performance increased on 12 of the 15 subordinate skills and decreased on 3 others. Analysis of covariance of the gain scores on the 32-item Subordinate Skills Post-Test showed the difference in performance to be significant and in favor of the group using the revised program ($F = 15.41$, $df = 1/100$, $p < .001$).

Interpretations

"The results of this experimental study show that adding instruction leading to improved performance on subordinate skills in a science learning program was successful in significantly improving performance on the criterion task. The practical implications for instructional development suggested by the findings are that . . . it is possible, by analyzing performance objectives . . . to determine probable subordinate skills to perform the criterion task. Following instruction, student performance on the subordinate skills can be used to pinpoint specific difficulties in the learning program."

A graph of the relationship of performance of all of the subjects on the Subordinate Skills Post-Test shows that the relationship to success on the criterion task is not linear. "Success on the criterion task appears to depend, to an increasing extent, upon the mastery of the complete set of subordinate skills." The authors point out that "although there was room for additional improvement, substantial progress was made with one revision of the learning program." They admit that the longer treatment received by the experimental group "violates one of the requirements of a research study" and identify this as a dilemma of the curriculum developer "who wishes to modify materials as he proceeds and still determine their effectiveness with the controlled conditions of a research design."

ABSTRACTOR'S ANALYSIS

This study should be of importance to designers of learning activity programs in that it supports the validity of using a learning model as a guide for the design and modification of learning activity programs to increase the learner's mastery of the desired learning outcomes. In justifying their use of Gagné's cumulative learning model, the authors state, "to derive a learning hierarchy for a task, the question is asked, 'What would the student have to know already in order to learn this task with a minimum of instruction'" (emphasis by abstractor). Yet, the program used in this study was modified by increasing the number of frames by 40 percent, a procedure that appears to be counter to "minimizing instruction."

The study avoids a major criticism of much educational research--the lack of a theoretical basis--by using a learning model as the rationale for analyzing and modifying a learning activity program to increase student attainment of the criterion task. Although the primary purpose of Gagné's subtask analysis has been identified (3) as that of maximizing the success of first-drafts of instructional materials, its use to analyze and modify existing materials to increase their efficiency of instruction appears to be both valid and useful.

Space limitations that are imposed on the writer of any journal article often leave unanswered many details and what may appear to be trivial questions regarding the research design, experimental procedure, data and data analysis. While the goal of the study was

"to investigate the prediction that a way to improve performance. . . is to provide additional instruction leading to attainment of subordinate skills failed by a substantial number of students," no criteria for success or failure was given. Nor is the definition of a "substantial number of students" provided.

The researchers fail to report whether or not the published program used in the study was designed using a cumulative learning model. Nor do they identify inadequacies in that program other than inadequate practice which must be inferred from the fact that 20 additional frames were added to expand the published program by 40 percent for the experimental group.

Gagné and Paradise (3) suggested that other possible explanations would be (a) that a subordinate learning set might have been omitted, and (b) that sufficient guidance in the subskills might not have been provided for efficient learning. An explicit statement that the analysis revealed no shortcomings in these respects would have added support to the experimental procedure of providing more review by adding frames rather than modifying frames that were judged to be inadequate or by improving upon their sequencing to provide a higher degree of mastery of the hierarchial subskills deemed necessary to master the criterion task. Did the original program designer assume too much with respect to the learner's mastery of the lowest tasks in the hierarchy? Was the original program optimally sequenced? These questions remain unanswered by the journal article.

It may be assumed that the goal of an optimal program is efficiency of learning which may be defined as mastering the criterion task in a minimal amount of time or with a minimal effort on the part of the learner, or both. Also, it may be inferred that the students in the experimental group required more time to complete a program expanded by 40 percent than did the students using the shorter published program. If so, it would have been helpful for the reader to know the mean times required by each group in order to assess more adequately the "increased efficiency" of the modified program.

A mean criterion test score of 3.84 out of a possible 7.0 was reported for students completing the modified program. While a mean gain of 0.82 points (12 percent) on the criterion test was found to be statistically significant, a 55 percent mean success level does not appear to represent a particularly high level of mastery for an "improved" learning activity program. Nowhere in the article is the mastery level defined explicitly. While mastery can, of course, be set at any level, evaluation specialists such as Mager, Gronlund, and Bloom and program writers commonly refer to mastery as being at the 80 percent level or higher. The relatively low level of mastery by both the control and experimental groups suggest that some of the students must have scored quite low on the criterion test and may not have "mastered" the criterion task--the goal of most learning activity programs.

In discussing the cumulative learning model, the authors say that "if the instructional materials are not successful in promoting mastery of the final task, a learning hierarchy has several built-in diagnostic features that . . . can be used to locate places of difficulty." While an increased performance on 12 of the 15 subskills is indicative of success in improving the efficiency of the program, what about the decreased performance on three (20 percent) of the subskills? Had additional frames been added to the published program to better develop these skills? If so, an analysis of the relevant parts of the modified program would have been in order. If not, is it possible that a single subtask analysis of the published program was inadequate? In either event, it would appear to have been desirable to have subjected the modified program to further subordinate skill analysis in order to increase further its effectiveness either by reducing instruction or by enabling students to achieve a higher mean level of achievement, or both.

The researchers do not state whether the tests and/or programs were administered by the classroom teachers or by the researchers. Nor do they indicate whether or not the students knew that they were involved in a study. While the journal article implies that the experimental group began to use the revised program as soon as the control group completed the Subskills and Criterion Post-Tests, it seems reasonable that several days might have been required to adequately analyze 37 items on two tests administered to the 49-student control group, identify weaknesses in the published program, write and edit 20 "new" frames, and reproduce the modified program for use by the experimental group. If so, this information should have been provided to the reader.

While the control and the experimental groups were provided different chemistry class activities during the study, might an interaction effect between friends assigned to each of the two treatment groups who studied together have influenced the post-test performance of students in the experimental group? Their study partners would have completed the published program. Study partners of students in the control group would have had no exposure to either program or, presumably, to solubility product problems.

The two treatment groups were comprised of students from only one school and only one learning activity program was used in the study. Thus, caution must be exercised in generalizing as to the effectiveness of using an "empirically-based subtask analysis" in the improvement of learning activity programs.

In summary, the study was described in a concise manner. A clear, description of the theoretically-based rationale for the study was provided. Details of the experimental procedure were limited by space restrictions imposed on a journal article. The written description is consistent with the data and the inferences are kept to a minimum.

Further research on the validity of using the cumulative learning model described by Gagné does seem warranted. While this study supports the hypothesis that additional practice on subordinate skills

does result in a significantly higher level of mastery of the criterion tasks, the hypothesis that "subordinate learning set may have been omitted" and that "insufficient guidance in the subskills had been provided" were apparently not tested. This study provides a substantial contribution to the initiation of more definitive studies of the implications of Gagné's model for the development of efficient learning programs.

REFERENCES

1. Gagné, Robert M. The Conditions of Learning, 2nd Ed. Holt, Rinehart & Winston, Inc., 1970.

2. Gagné, Robert M. and Leslie J. Briggs. Principles of Instructional Design. Holt, Rinehart & Winston, Inc., 1974.

3. Gagné, Robert M. and N. E. Paradise. "Abilities and Learning Sets in Knowledge Acquisition," Psychological Monographs: General and Applied, Vol. 75, No. 14. Whole No. 518, 1961.

4. Tyler, Ralph, R. M. Gagné, and Michael Scriven. Perspectives of Curriculum Evaluation. A.E.R.A. Monograph Series on Curriculum Evaluation, No. 1. Chicago: Rand McNally & Co., 1967.

EJ 027 041

Ritz, William C. and Ronald J. Raven, "Some Effects of Structured Science and Visual Perception Instruction Among Kindergarten Children." Journal of Research in Science Teaching, Vol. 7, No. 3:179-186, 1970.

Descriptors--*Achievement, *Elementary School Science, *Instruction, Kindergarten Children, *Reading Readiness, Science Education, *Visual Perception

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

Purpose

The purpose of the study was to investigate the effects of two instructional programs (Science - A Process Approach, Part A, and the Frostig Program for the Development of Visual Perception) on the attainment of reading readiness, visual perception, and science process skills in kindergarten children. Specifically, the investigators were concerned with answering the following questions:

(1) What is the effect of perceptual training given at the conclusion of science instruction utilizing one of the new elementary science programs?

(2) What effect does instruction using one of the new elementary science programs have on perceptual abilities?

(3) Does science instruction involving "hands-on" experiences enhance reading readiness skills?

(4) What effect does a combination of science instruction involving "hands-on" activities and perceptual training have on reading readiness skills?

Rationale

Recent research indicates that some of the new innovative elementary science programs (e.g., SAPA, SCIS, etc.) do enhance the reading readiness skills of kindergarten pupils. Thus, these programs may indeed be a starting point for developing language readiness in preschool children. The data reported by Ayers and Mason (1) appear to support this position.

Since training in visual perception may play a major role in the development of reading readiness attainment, it is this aspect of the activities in Level A of the SAPA program that may indeed make the greatest contribution. The Frostig Program for the Development of Visual Perception is also designed specifically for this purpose. The investigation reported by Raven and Strubing (3) suggest and support this relationship between the Frostig program and SAPA. Thus,

instruction in either one or a combination of both programs should contribute to and enhance reading readiness through visual perceptual training.

Research Design and Procedure

Twenty-four kindergarten classes (n = 192) in nine different school systems in Central and Western New York State and one in Northwestern Pennsylvania were randomly assigned to three treatment groups. Since most of the kindergarten teachers were assigned two classes per day, the toss of a coin determined which class would participate in the study. Six teachers each in two of the treatment groups were already using SAPA in the primary grades of their schools. The remaining four teachers, two each in the same treatment groups, were using SAPA, Part A, for the first time. None of the teachers in the two treatment groups had ever used the Frostig materials in their kindergarten classes prior to the study. There were eight classes and eight teachers in the third group. The study was conducted over a period of seven months.

Group I received instruction in SAPA, Level A, exercises a through k, and no instruction in the Frostig "Perceptual Constancy" unit. Group II received instruction in SAPA exercises a through k, and instruction in the Frostig "Perceptual Constancy" unit involving worksheets 1-14 and 38-48 at the conclusion of the science exercises. Group III received no instruction in the SAPA exercises and then later received instruction in the same Frostig exercises as Group II.

All three groups were tested at the completion of exercises a-k of SAPA, Level A (Instructional Phase A), and again at the conclusion of the Frostig "Perceptual Constancy" unit exercises (Instructional Phase B). The Metropolitan Readiness Test, Form B, was used to collect data on reading readiness attainment and was administered by the kindergarten teachers. The Frostig Developmental Test of Visual Perception was used to collect data on visual perception attainment and the Competency Measures for Groups instrument was used to collect data on science process attainment. The latter two instruments, because of prior training requirements, were administered by ten outside test administrators of whom all but one were ex-teachers.

To test the hypotheses of no significant differences, a one-way analysis of variance (ANOVA) was computed for the mean scores obtained on the above instruments.

Findings

At the conclusion of the first Instructional Phase (A), no significant differences were found between the group mean scores on the Metropolitan Readiness Tests (Form B), and the Frostig Developmental Test of Visual Perception (Perceptual Quotient and Scale Score, Subtest III). However, a significant F value for science scores as measured by the Competency Measures for Groups was recorded. Thus, the pupils in the two groups who participated in the SAPA exercises

(Groups I and II) did score higher in science process attainment than did the pupils (Group III not using the SAPA science program ($p < .025$).

No significant differences in test mean scores were found at the conclusion of Instructional Phase B with respect to readiness skills and science process scores. However, there was a significant F value for both Perceptual Quotient and the Frostig subtest III mean scores. Thus, the hypotheses of no treatment differences among groups with respect to visual perception skills at the end of Instructional Phase B was rejected. Pupils (Groups II and III) receiving instruction in perceptual training did score higher than pupils (Group I) not receiving perceptual instruction ($p < .01$).

Interpretations

It is difficult to draw conclusions from the study as reported. Such limitations as time, money, and the fact that only parts of the programs were used make it difficult to assess the value of the complete programs as the investigators note. They also write that the "complexity of 'reading readiness' also contributes to producing what may be a misleading representation of the middle ground which has been postulated to exist between SAPA (Part A), visual perception, and reading readiness."

The facts that participation in the SAPA exercises did not result in significant scores on the Metropolitan Readiness Tests, Form B, and that participation in the Frostig Perceptual Unit did not result in significant readiness scores seem to contradict results of the Ayers and Mason study. Thus, participation in one of the new and innovative science programs such as SAPA and perceptual training have little or no significant effects upon reading readiness skills. But, at the same time, performance scores revealed that the programs used did not detract or have a negative effect on reading readiness attainment. This finding serves as evidence and supports the contention that significant time periods can be devoted to science education in the elementary school program without detracting from other important educational outcomes such as "readiness."

Although there was a significant improvement in visual perceptual abilities and science process skills, further investigation is required to determine the possible effects of Science - A Process Approach exercises on "reading readiness." In order to do this and obtain more valid results, the complete SAPA, Level A, exercises should be used. In addition, two different and diverse measures of reading readiness should be used in the study. Further, science process skills, if assessed, should involve individual testing because this may be "the most reliable method of assessing the attainment of the behavioral skills of Science - A Process Approach." Along these same lines, individual assessment of science process skills coupled with the use of the Competency Measures for Groups instrument might be explored. The value for this procedure would reduce by sixty percent the total number of individual competency measure tasks which would have to be

administered in order to assess attainment of the behavioral skills of Science - A Process Approach at the kindergarten level.

ABSTRACTOR'S ANALYSIS

A number of studies similar to the present study have been conducted within recent years involving some of the new and innovative elementary science programs (e.g., SAPA, SCIS, ESS, etc.) which stress the "hands-on" approach. Generally speaking, these studies are attempts to manipulate early experiences by providing "hands-on" activities with objects and materials common to the environment of the children under study. They are generally divided into deprivation or enrichment studies for the most part. Much of the research was generated during the 1960's and usually dealt with the effects of some kind of educational intervention in the lives of preschool children living in poverty areas. It is difficult to determine into which category this study falls. This is mainly due to the fact that the investigators did not describe the population from which the sample was chosen. Replication of the study is impossible. The elementary school program is not described and thus it is impossible to determine whether or not it is an intervention or enrichment study. One can assume that it is probably the latter case, but this is only an inference.

The results of the study shed little new information on the concept of "readiness" and the contributions of perceptual training. It is well established that perceptual skills are at least a necessary condition for the acquisition of reading skills. The research suggests that perceptual skills are involved in reading success as well as in reading failure. Children who are good readers appear to have well-developed perceptual skills while poor readers appear to be deficient in this regard. But this aspect is not considered in this report and thus the question is moot here.

Readiness is a rather complex concept and is not fully understood nor completely translatable into elementary instructional programs presently. However, auditory perception in recent years or the discrimination of speech sounds is important for learning to read. Indeed, the degree and quality of oral language the preschooler brings to bear upon receiving reading instruction is also an important factor. Thus, many factors are bound-up in the concept of "reading readiness." This study has only looked at a small portion of the complex concept and has offered little in the way of new insights or information.

The research design of the study is straightforward. However, the investigators refer us to Figure 1 on page 181 for the paradigm. Nowhere does this appear in the report. Instead, the reader is presented with a hierarchy chart entitled "General Hierarchy" which is never referred to in the body of the report. Perhaps this is an oversight on the part of the investigators or the publishers. But it is most confusing since the report refers to the figure several times.

Some background research is cited in the article, but not enough. The summaries of the research reported are very sketchy and thus

limited in value to the reader. In writing the discussion, the investigators only refer to one of two of the research studies cited in the introduction. It would have been much more reasonable to quote similar studies recently conducted in order to place the study in perspective with current research efforts. However, this is left up to the reader and, with non-significant results, one is hard pressed to really evaluate the true worth of the study or its significance to current classroom practices and cognitive developmental research.

As stated previously, a total lack of sample description negates total replication although little would be gained from such an effort. However, if the stated objective of the study were tried out on middle-class SES children, the the results would have been limited to begin with because of possible home environmental factors. If the children were from a low SES, then the results are contrary to many preschool intervention programs which are being conducted nationally (e.g., Head Start, Project Literacy, etc.). This is truly disturbing because the reader cannot reach any true and valid conclusions. Research is not meant for this purpose.

It was noted that to insure uniformity of scoring, a team of two test administrators scored all Frostig papers. Nowhere is there mention of determining whether or not there is agreement between scorers. An interobserver reliability coefficient should have been computed by randomly selecting some of the scored tests and plugging in the results into the simple formula. The scoring done in the above manner can introduce scoring bias. The same case applies to the remainder of the tests which were scored by only one of the test administrators. Again, there is the possibility of scoring bias which can indeed invalidate the study or at least place the results in question. Again, an interobserver reliability coefficient should have been computed, thus eliminating a major source of possible bias in the study.

Research has suggested that indeed there may be sex differences in readiness skills. It is an established fact that girls read generally earlier than boys. Therefore, it might have proven significant to this study if the investigators had looked at the data to see if indeed this situation may have existed. It is perhaps something that the investigators can do by reanalyzing the data. This might prove to be of significance, particularly in relation to science instruction and activities and their differential effect on readiness skills in girls and boys.

The investigators did acknowledge that the study involves "repeated observations of the same subjects." Univariate analysis of variance (ANOVA) F tests were computed for the scores obtained on each of the instruments used in the study. To use this statistical procedure, the assumptions of normality and homogeneity of variance within cells must be met. In addition, Winer (4) stresses that in using this procedure, the additional assumption of homogeneity of covariances (i.e., correlation) between repeated measures must be met. Violation of these assumptions can result in too many rejections of the null hypothesis for the stated level of significance [McCall and

and Appelbaum, (2)]. If indeed this situation does occur, then multivariate statistical procedures may be necessary. When the n is large, multivariate analysis is generally the best choice in most cases.

Additional studies are required to get at specifically what aspects of elementary science education can enhance "readiness" skills and to what degree. How much exposure to science instruction and the quality of science instruction must be determined. By necessity, these studies may have to be longitudinal in nature. This will require large research budgets and large numbers of trained personnel working in the field (i.e., the school setting). This by design will probably eliminate the one-shot standard dissertation study type. However, careful planning and close supervision can certainly utilize graduate student personnel with resulting data sufficient for dissertation research reports. This, however, should not be the ultimate goal of such studies.

Certainly, more studies need to be conducted to determine the sex differences in "readiness" skills. Why is there a difference and what factors contribute to its manifestation under control conditions? What effect does science instruction have on "readiness" skills in girls and boys and where are the differences found, if any? These are questions that science educators must address themselves in future research.

REFERENCES

1. Ayers, Jerry B. and George E. Mason, "Differential Effects of Science - A Process Approach Upon Change in Metropolitan Readiness Test Scores Among Kindergarten Children," The Reading Teacher, 22:435-439, February, 1969.

2. McCall, Robert B. and Mark E. Appelbaum, "Bias in the Analysis of Repeated-Measures Designs: Some Alternative Approaches," Child Development, Vol. 44:401-415, 1973.

3. Raven, Ronald and Herbert Strubing, "The Effect of Visual Perception Units on Achievement in a Science Unit: Aptitudinal and Substantive Transfer in Second Grade Children," American Educational Research Journal, Vol. 5, No. 3, 1968.

4. Winer, B. J. Statistical Principles in Experimental Design (2nd edition). New York: McGraw-Hill, 1971.