DOCUMENT RESUME

ED 245 942 SE 044 676

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

TITLE Investigations in Science Education. Volume 10, Number 2.

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

PUB DATE 84

NOTE 75p.

AVAILABLE FROM Information Reference Center (ERIC/IRC), The Ohio State Univ., 1200 Chambers Rd., 3rd Floor, Columbus, OH 43212 (subscription $8.00, $2.75 single copy).

PUB TYPE Information Analyses (070) -- Collected Works - Serials (022) -- Guides - Non-Classroom Use (055)

JOURNAL CIT Investigations in Science Education; v10 n2 1984

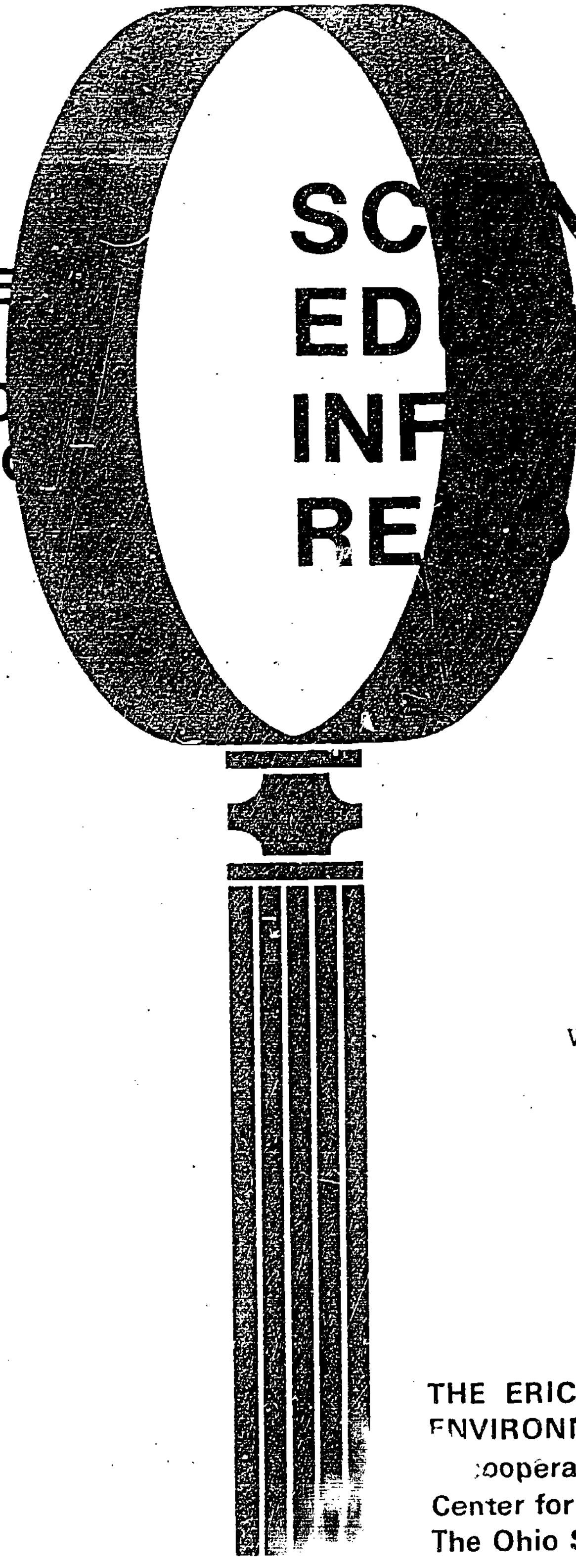
EDRS PRICE MF01/PC03 Plus Postage.

DESCRIPTORS Academic Achievement; Computer Assisted Testing; Elementary Secondary Education; Environmental Education; Higher Education; *Inservice Teacher Education; *Preservice Teacher Education; Questioning Techniques; Science Education; *Science Instruction; Science Tests; *Teacher Attitudes; *Teacher Behavior; Time Factors (Learning)

IDENTIFIERS *Science Education Research; *Wait Time

ABSTRACT

Presented are abstracts and abstractors' analyses of 10 studies related to one or more aspects of teacher education and 2 studies in a "miscellaneous" category. Analyses in the first section (teacher education) are on studies of: the use of wait-time and its effect on science achievement; teacher competencies; the influence of teacher behavior on student performance; affective predictors on preservice science teaching behavior; ideal teacher behavior perceptions of science students; the effectiveness of training methods in modifying questioning and wait time behaviors of Thai high school chemistry teachers; three methods of improving preservice science teachers' questioning knowledge and attitude toward questioning; the effectiveness of a basic science skills course for preservice elementary teachers; the development of a test to measure teachers' conceptions of the meaning of science; and teacher education majors compared to other majors relative to several variables in an attempt to promote positive attitudes toward science and science instruction. Analyses in the second section ("miscellaneous") are on studies of: the effects of frequent multiple-choice testing with immediate computer feedback and the assessment of the universality of participation in pro-environmental behavior. (JN)



SCIENCE EDUCATION INFORMATION REPORT

INVESTIGATIONS IN
SCIENCE EDUCATION

Volume 10, Number 2, 1984

THE ERIC SCIENCE, MATHEMATICS AND
ENVIRONMENTAL EDUCATION CLEARINGHOUSE
in cooperation with
Center for Science and Mathematics Education
The Ohio State University

INVESTIGATIONS IN SCIENCE EDUCATION

National Association for Research in Science Teaching

Clearinghouse for Science, Mathematics
and Environmental Education

INVESTIGATIONS IN
SCIENCE EDUCATION

Volume 10, Number 2, 1984

Published Quarterly by

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

Subscription Price: $8.00 [illegible] Single Copy Price: $2.75. Add [illegible].00
for Canadian mailings and [illegible] foreign mailings.

Volume 10, Number 2, 1984

NOTES FROM THE EDITOR:

Ten of the twelve articles critiqued in issue two of Volume 10 of Investigations in Science Education relate to one or more aspects of teacher education. Tobin investigated the use of wait-time and its effect on science achievement. Chiappetta and Collette studied teacher competencies. Shymansky and Penick commented on the influence of teacher behavior on student performance. Sunal examined affective predictors on preservice science teaching behavior. Peterson and Mayes investigated ideal teacher behavior perceptions of science students. Chewprecha et al studied the effectiveness of training methods for modifying teacher questioning behavior, as did Riley in a different study. Sherwood and Gabel described the effectiveness of a basic science skills course for preservice elementary teachers. Cotham and Smith reported the development of a test to measure teachers' conceptions of the meaning of science. Gabel compared teacher-education majors with other majors relative to several variables in an attempt to promote positive attitudes toward science and the teaching of science.

In the "miscellaneous" section are to be found critiques of articles on the effects of frequent multiple-choice testing with immediate computer feedback (Fisher et al.) and the assessment of the universality of participation in pro-environmental behavior (Larson et el.).

Patricia E. Blosser
Editor

Stanley L. Helgeson
Associate Editor

TEACHER

EDUCATION

Tobin, Kenneth G. "The Effect of an Extended Teacher Wait-Time on Science Achievement." Journal of Research in Science Teaching, 17 (5): 469-475, 1980.

Descriptors--*Academic Achievement; *Middle Schools; *Questioning Techniques; Science Education; *Science Instruction; Secondary Education; Secondary School Science; *Teacher Behavior

Expanded abstract and analysis prepared especially for I.S.E. by Dorothy L. Gabel, Indiana University.

Purpose

The author's purpose for conducting this study was to determine whether the use of an extended period of time (wait-time) preceding any teacher utterance had an effect on the science achievement of middle school students.

Rationale

Studies by Rowe (1974) and Lake (1973) determined that changes in wait-time have an effect on student behavior. Beh[illegible] [illegible]anges that occur due to increased wait-time appear to increase s[illegible]dents' participation in the instruction and may have a positive effect on achievement. In this study, Tobin extends the work of Rowe and Lake by determining the effect of increasing wait-time on student achievement in science.

Research Design and Procedures

The sample consisted of 23 intact classes of 11-13 year old students in 11 Australian middle schools. Because it was impossible to randomly assign teachers to the treatment and control groups, teachers within given schools were paired with those in other schools according to numbers of teachers per school and teaching experience. Schools were then randomly assigned to treatment and control groups.

The study consisted of three phases. During the first phase all teachers used their normal wait-times during instruction. During phase 2, the treatment group teachers tried to increase their wait-time. Phase 3 was a replication of phase 2 using a different science unit.

The design as given in the report is as follows:

$$\frac{X_1 O_1 X_2 O_2 X_2 O_3 O_4}{X_1 O_1 X_1 O_2 X_1 O_3 O_4}$$

where X_1 = normal wait-time used in instruction
X_2 = extended wait-time used in instruction
O_1 = summative measure (Ice Cubes)
O_2 = summative measure (Colored Solutions)
O_3 = summative measure (Clay Boats)

O_4 was not defined by the author.

Each lesson was recorded on a portable tape recorder attached to the teacher's waist. At a later time the duration of pauses was measured with a servochart plotter. An estimate of the mean teacher wait-time was obtained for each teacher by averaging a sample of approximately 50 pauses randomly selected from the tapes of each lesson. The experiment lasted for 13 weeks and included 7 lessons for phase 1, 6 lessons for phase 2, and 7 lessons for phase 3.

The achievement tests for each of the three phases included items covering the range of levels of Bloom's taxonomy, a large proportion being application level or higher. Reliabilities of the tests as determined using the KR-20 formula were 0.6, 0.6, and 0.7. Data were analyzed using multiple regression techniques with wait-time and achievement on the phase 1 test as independent measures.

Findings

In phase 1 where teachers used their normal wait-time, the mean wait-time for all teachers was 0.5 seconds. The correlation between mean wait-time and mean class achievement was close to zero indicating that no relationship between wait-time and science achievement existed in this phase of the study.

During the experimental phases of the study, the 10 teachers in the normal wait-time group maintained an average of 0.7 seconds while the 13 teachers in the extended wait-time group averaged 3.1 seconds.

Results of the multiple regression analyses indicated that the mean wait-time was significantly related to achievement on Test 3, but not significantly related to achievement on Test 2.

Interpretations

The results of this study support the hypothesis that the use of an extended wait-time will lead to higher science achievement. Tobin attributes the lack of significant findings for Test 2 as possibly due to the fact that it may take both teachers and pupils a period of time to adapt to extended wait-times.

ABSTRACTOR'S ANALYSIS

The major strength of this study is in its excellent design. Tobin overcame the problem of being unable to assign teachers to the treatment and control groups randomly by pairing schools and then randomly assigning one school of the pair to the treatment group, the other to the control group. The taping of the lessons and the use of the servo/plotter gave excellent control of one of the independent variables, wait-time. To show that the three units were related so that generalizations could be made, correlation coefficients were calculated. This was also a strong point in the design of the study.

As in most any study, the design could be improved given the finances and the time. A major drawback in this study is the small number of teachers in the sample. A sample size of 23 is questionable for the use of multiple regression techniques with two independent variables. If the study were divided according to only two levels of the independent variables, the size of some cells would be only five. A larger sample size would have increased the credibility of the findings considerably.

Some minor omissions in the reporting of the study lessened its excellence. In the design, an O_4 was listed but never defined. One wonders why this was included. No mention was made of the number of items in each of the three tests used as dependent measures. The rather low reliability coefficients given for these tests (0.6 to 0.7) might be attributed to a small number of items, but there is no way of knowing this. Mention was made of the fact that these tests were examined for reading level and content validity by a panel of three judges. However, the level of agreement among the judges was not given.

Tobin did a creditable job in interpreting the data. However, a slight modification of the analyses might have strengthened his conclusions. He states that the reason for lack of statistical findings for phase 2 is probably attributed to the fact that it takes time for teachers and students to adjust to extended wait-times. Tobin states in his report that 5 of the 13 teachers in the extended wait-time group failed to attain a 3 second average wait-time. Including the data from these five teachers certainly would dilute the results. Perhaps data from these teachers should have been eliminated, and the data reanalyzed. Since these teachers' data were included in the analyses, the results of the study might better be expressed as "teachers encouraged to use an extended wait-time in teaching science produce higher achievement than teachers who are not encouraged to do so" rather than as "the use of an extended teacher wait-time will lead to higher achievement."

Tobin also gives in the report the combined average wait-time for the two experimental phases of the study. Had these been listed separately, one might be able to determine if teachers were adjusting to the extended wait-time by an increased wait-time for phase 3 over phase 2. If the wait-time did not increase for phase 3, it would appear that the teachers had adjusted during phase 2 and Tobin's interpretation of the findings would not be credible. Perhaps wait-time is more critical alternative explanation topics than others. This would be a credible alternative explanation of the results. By listing the average wait-times separately for each phase of the study, Tobin might have even been able to make a generalization about the effectiveness of various extended wait-times. Since the publication of this report, studies of this nature have been done and reported in the literature.

REFERENCES

Lake, J. H. "The Influence of Wait-Time on the Verbal Dimension of Student Inquiry Behavior." Doctoral Dissertation, Rutgers University, 1973. Dissertation Abstracts International, 34 6476-A, 1974.

Rowe, M. B. "Wait-Time and Rewards as Instructional Variables, Their Influence on Language, Logic and Fate Control: Part One - Wait-Time." Journal of Research in Science Teaching, 11 263-279, 1974.

Chiapetta, Eugene L. and Alfred T. Collette. "Identification of Science Teacher Competencies for Implementing ISIS Minicourse Instruction." Science Education, 64 (1): 53-58, 1980.

Descriptors--*Educational Research; Integrated Curriculum; *ISIS Instruction; *Minicourses; Research Methodology; *Science Course Improvement Projects; *Science Education; *Secondary School Science; Secondary Education; *Science Teacher Competencies; Secondary School Students; *Teacher Characteristics; *Teacher Effectiveness

Expanded abstract and analysis prepared especially for I.S.E. by Gerald H. Krockover, Purdue University.

Purpose

The purpose of this study was to determine practitioner perceptions of the competencies that science teachers should possess to implement ISIS minicourse instruction.

Rationale

Teacher competence is a significant factor in the success of any program. Many questions can be raised relative to the success of the Individualized Science Instructional System (ISIS) in the nation's schools. For example, will state departments of education and school districts adopt soft-bound textual material? Will they subscribe to minicourse or modular material? Is there enough pure science in the ISIS units for high school science programs? And, can science teachers manage self-paced instruction? This study deals with the last question raised: that is, can science teachers manage self-paced instruction?

Research Design and Procedure

The population for this study consisted of 70 science teachers field testing ISIS minicourses at eight funded trial centers during the 1975-1976 school year. Procedures used in this study were adopted from the Delphi method. The research was accomplished using a three-round procedure. In the first round competency statements were generated by the participants. In the second round the participants selected levels on the Taxonomy of Cognitive Objectives (Bloom, 1956) at which they felt each competency statement should be specified. In the third round the participants rank-ordered the competencies in terms of their importance. The rate of usable returns received from the ISIS trial teachers in rounds one, two and three was 61, 47 and 58 percent respectively. In round one the ISIS trial teachers were asked to list those competencies (skills, knowledge, and attitudes) that all science teachers who implement minicourse instruction should possess. The responding teachers provided 230 statements. The responses were given to a panel of five judges and they were instructed to sort all responses into common categories and to construct a label for each category. The judges grouped the responses into three categories: cognitive, affective, and personality. They identified seven personality characteristics, four affective competencies, and thirteen cognitive competencies. Since each cognitive competency consisted of one or more competency statements, rounds two and three were devoted to the cognitive competencies.

In round two the respondents were asked to assist in a process for adding greater specificity to the competency statements associated with the cognitive competencies. They were directed to identify a level on the cognitive taxonomy at which science teachers implementing minicourses should be able to demonstrate a given skill. One-sample X^2 tests were used to determine taxonomic levels consistently chosen by the respondents to specify the competency statements. Competency statements with associated X^2 values significant at $p \leq 0.05$ were written at the specified levels.

In round three the respondents were asked to rank-order the cognitive competencies. Each competency consisted of a title and one or more competency statements that defined the competencies relative to one of the taxonomic levels.

Findings

Fourteen competency statements showed a significant difference, $p \leq 0.05$, in the analysis, while five did not. The elimination of five competency statements for which no consensus was achieved consequently reduced the number of cognitive competencies from 14 to 9. Round three participants ranked the nine cognitive competencies in order of importance. Rank-order was determined by the sums of ranks assigned to each competency. Competencies one to five were seen by respondents as considerably more important than six to nine.

The Kendall Coefficient of Concordance (W) was computed to determine the degree of consistency of the ranking of the nine competencies. A concordance value of 0.12 ($p \leq 0.001$) was found for the ranking. This value suggests a rather low degree of consistency among the ranking of the nine competencies. Therefore little confidence can be placed on the exact order of importance of the competencies as rated by the trial teachers.

The five competencies that were judged most important to the trial teachers included: using a variety of instructional strategies, promoting individualized instruction, showing acceptance and respect for students, need to control the classroom, and organizing the classroom to facilitate instruction.

ABSTRACTOR'S ANALYSIS

This study made a noble attempt to identify the science teacher competencies needed for implementing ISIS minicourse instruction. However, a serious flaw in the study is the poor response received from the ISIS trial teachers in rounds one, two and three. It is unfortunate that only 61% of the trial teachers responded in round one, 47% in round two, and 56% in round three. A much higher degree of response would have greatly assisted in the credibility of this study. Furthermore, while other studies have shown the importance of affective competencies and personality characteristics in becoming a successful science teacher, this study chose to ignore those two areas and concentrate on the cognitive area instead. The top five cognitive competencies indicated by the teachers: using a variety of instructional strategies, promoting individualized instruction, providing a humanistic learning environment, controlling the classroom, and organizing the classroom to facilitate instruction would apply to any "good" science teacher regardless of the program being used. It is unfortunate that the research study was unable to capture the unique characteristics of the ISIS materials and program. As a result the questions asked in the introduction to this study were not dealt with accordingly.

The final statement in the study indicates the need for searching out the specific characteristics and attributes of the ISIS materials since the uniqueness of the ISIS program is critical to the success of the science teacher. For, as the authors state, "The ISIS teacher must possess and radiate a desire to work with young people who experience difficulty in their work, and to spend the time to gather the necessary equipment and materials for self-directed learning. These skills and attitudes probably are found among science teachers who are regarded by colleagues as being creative, enthusiastic, open-minded, patient, and self-confident." Isn't this what all teaching is about?

Shymansky, J. and J. Penick. "Teacher Behavior Does Make a Difference in Hands-On Science Classrooms." School Science and Mathematics, 81 (5): 412-422, 1981.

Descriptors--Academic Achievement; *Creativity; Elementary School Science; *Elementary Secondary Education; *Inquiry; Science Course Improvement Projects; Science Education; *Science Instruction; Secondary School Science; Student Characteristics; *Student Improvement; *Teacher Behavior; *Teaching Methods; Teaching Styles

Expanded abstract and analysis prepared especially for I.S.E. by David P. Butts, University of Georgia.

Purpose

The purpose of this study was to describe the evidence of linkage between teaching behavior and student performance in science classrooms. This involved ascertaining how much teacher involvement and what type of involvement helps students learn in a science classroom that emphasizes hands-on activity.

Rationale and Procedure

Within a contrasting paradigm of teacher dominated instruction and student structured instruction, this report identified findings from five research studies in which the same paradigm was used.

Findings

The findings of these studies showed a pattern of outcomes, namely:

One -- the teacher dominated strategy resulted in students being dependent on the teacher,

Two -- students view science and scientists more positively in a student centered environment,

Three -- students show more on-task behavior in student centered environment, Four -- student creativity and problem solving is higher in student centered environments.

Interpretations

In general the authors concluded that if the goal of science instruction is to help students solve problems and do this creatively, the teacher who uses a student centered strategy will obtain superior results. If the teacher wished to manage counter-productive student classroom behavior, this can be done by reducing the teacher's restrictive classroom behavior.

ABSTRACTOR'S ANALYSIS

This synthesis of five research studies around a single question of interest to science teachers is a positive contribution. The synthesis could have been strengthened if the authors had provided more information about the samples, topics taught, and grade level of the subjects. As in meta-analysis studies, this report does add support to ways in which teachers can make a significant difference in student classroom behavior and outcomes. That these results seem to be consistent over several studies certainly strengthens the validity of the conclusions. It would help the reader, however, to have had more of the specific data on which the conclusions are based.

Sunal, D. W. "Affective Predictors of Preservice Science Teaching Behavior." Journal of Research in Science Teaching, 19 (2): 167-175, 1982.

Descriptors--*Affective Behavior; *Affective Measures; Elementary Education; Elementary School Teachers; Formative Evaluation; Higher Education; *Predictor Variables; *Preservice Teacher Education; Science Education; *Science Instruction; Teacher Attitudes; *Teacher Behavior; Teacher Characteristics

Expanded abstract and analysis prepared especially for I.S.E. by Willis Horak, The University of Arizona.

Purpose

This study was designed to analyze the affective changes in elementary pre-service teachers at three specific times during their teacher education program. These times were chosen to correspond to the periods when students are perceived as developing and refining their basic classroom beliefs and behaviors. The study measured and reported the attitudes toward teaching science, the attitudes toward children learning science, and the attitudes toward nature. The pre-service teachers' stated attitudes were also related to actual classroom practices through an observational [illegible] ing system used during the field based elementary science methods class and during the student teaching experience. A major part of the study was the instrument development and validation.

Rationale

The rationale for this study centered around the need to describe and analyze the factors which contribute to the unmatched role performance of teachers and the curriculum role needs. The modern elementary science curricula generally emphasize activity oriented science. The role of the teacher is thus one of a helper and guide. They are required to exhibit the ability to use basic skills of scientific investigation and to help children initiate and implement their own methods of problem solving. However, as reported in the literature, this is often not the case. Elementary science teachers often engage in a large amount of dominating classroom behavior and also require a large amount of student recitation.

Affective variables were chosen to be measured and analyzed. The previous research has indicated that a prime aspect of the successful implementation of new curricula lies in the acceptance of the proposed teaching beliefs and behaviors. The new elementary science curricula generally have very little impact if they are adopted, the content covered, but the teaching behaviors unchanged. It was felt that the affective factors identified in this study could be used subsequently by instructors during the planning of field based science methods courses.

Research Design and Procedures

Four instruments were used to measure the affective and the teaching behavior attributes of a sample of thirty pre-service teachers. Approximately 96% of the participants were female. Additionally three-fourths of the population came from the same geographic area. The attributes were measured on three separate occasions. These were (1) at the beginning of the science methods course, (2) at the end of the science methods course, and (3) at the end of the student teaching experience. Two affective instruments were designed specifically for this study. Two additional instruments utilizing observational rating systems had been previously used and described.

Interest toward science was measured with a Teacher Preference Scale. This instrument consists of items relating to a teacher's interest in and preference for teaching specific subjects. A high mean score indicates the expressed preference for teaching science.

Attitudes toward science were measured with the Semantic Differential Instrument for Science Teaching. This instrument is divided into three sub-tests. One measures concepts related to nature. A second section measures attitudes related to elementary school children learning about science. The third section measures beliefs about myself teaching science to children. The instrument itself was further analyzed by utilizing factor analysis procedures.

Findings

The data analysis related to both detecting significant changes in pre-service teachers' beliefs and to predicting their teaching behavior based upon these changes. The SDIS Instrument was submitted to a factor analysis. Thirteen specific factors were identified among the three sub-parts of the instrument. Means and standard deviations were calculated and reported on each of these factors for the three specific time periods. Analysis of variance procedures were used to identify significant changes in teachers' attitudes over time.

This analysis revealed significant changes in attitude for five of the thirteen factors. These factors were comfort as related to the concept of nature, interest and security as related to the concept of children learning science, and stimulation and safety as related to myself teaching science. There was also a significant change in the scores on the Teacher Preference Scale.

The scores on the five factors identified as significantly changed were then used to predict or account for differences in observed teaching behavior as measured by the Micro-teaching Skills in Science Checklist (MSS) and the Survey of Classroom Activities in Science (SOCAS). This was done utilizing regression analysis techniques. The MSS was administered during the science methods course and the SOCAS was administered during student teaching.

Seventy percent of the variance of the scores on the MSS instrument were accounted for by the six affective factors. On the subscales of the MSS, 74 percent of the variance of lesson planning behavior scores and 64 percent of the variance of the teaching behavior scores were accounted for by the six affective factors.

When the scores on the SOCAS were analyzed using regression techniques, 53 percent of the variance was accounted for by the affective measures. On the subscales of this instrument 55 percent of the variance of traditional teaching behaviors, 33 percent of the variance of performance of teacher inquiry behaviors, and 52 percent of the variance of student performance in inquiry skill behaviors were explained.

Interpretations

This study pointed out that significant affective changes do take place during a field based elementary science methods course. The changes in beliefs and attitudes were also found to be useful predictors of actual teaching behavior. The specific areas of interest in teaching science, attitude toward teaching science, and beliefs about children learning science should be measured early in the elementary science methods course. The results of the measures could be used to more effectively individualize instruction in these courses. Measures of attitudes toward nature do not appear to effectively predict teaching behavior. It is therefore not reasonable to use such measures in an elementary science methods course.

ABSTRACTOR'S ANALYSIS

The research reported in this article is of the type that should be a main concern of most elementary science methods instructors. Too often we assume that the expressed beliefs and attitudes of our students are directly related to their subsequent teaching behaviors. By observing them during their student teaching experience, we find that this is often not the case. There is a wide variation in the teaching behaviors which are being modeled by the student teachers. Many times it also appears that a pre-service student is apparently saying one thing and doing something entirely different. This study was influential in identifying which affective variables are worthy of consideration if we are concerned with affective variables predicting teaching behaviors.

The analysis of the data is very well reported. It is always useful for researchers to see the actual means and standard deviations of the variables measured. It would be refreshing if more journals and researchers consistently reported the data. The simple reporting of only significant differences often does not entirely explain the situation.

For this study, it would probably been helpful for the researcher to make some attempt to analyze the teaching behavior of the involved cooperating teacher. I realize this was not the overall intent of the study. However it is sometimes hard for elementary student teachers to model behaviors which their cooperating teachers do not value. Often if the cooperating teacher does not engage in a lot of inquiry-oriented science, the student teacher will not, either. Often the physical set-up of the room makes activity-oriented science hard to implement. A follow up to this reported study might possibly involve the congruence of student teacher and cooperating teacher belief systems.

The factor analysis of the Semantic Differential Instrument for Science Teaching needs to be extended. The sample size of 148 used for the factor analysis procedure is relatively small. I generally like to see larger sample sizes used when factor analysis procedures are being employed in data analysis. It will be interesting to see if the identified factors remain stable when additional scores are added to the presently accumulated data. The nature concept subscale reported factor structure is puzzling to me. During the description of the instrument development, it is stated that each concept is rated on eighteen adjectival pairs. However Table I reports nineteen adjective pairs loaded on the nature concept. The author needs to check the accuracy of this table.

The Teaching Preference Scale could be described more in depth. It would be useful to be given an example of the types of activities described by the six items. From the information given, this does not appear to be a very involved instrument. Is science coordinated with the teaching of mathematics, and the teaching of social studies? Is mathematics ever paired with language arts or must the student choose between science and something else? Do the six items compare the curricular areas on teaching behaviors or on subject matter content? From the report one cannot tell.

In summary, I feel that this is a well done research study. First of all, the topic is relevant to anyone engaged in science teacher education. The information gained should be quite useful for planning effective courses. The criticisms of the study are minor. Most often they are the result of not discussing enough. However anyone interested in the study could readily get the information from the author. It would be useful to compare the field based elementary science methods course with a non-field based elementary science methods course. This needs to be done as a follow up to the reported study. Once we know what affective factors apparently influence teaching behaviors, it becomes necessary to develop diverse teaching modules which successfully change attitudes and beliefs of our pre-service teachers. This study can thus serve as a focal point of many needed follow up studies.

Peterson, Kenneth and Bea Mayes. "Ideal Teacher Behavior Perceptions of Science Students: Success, Gender, Course." School Science and Mathematics, 81 (4): 315-321, 1981.

Descriptors--Attitudes; *Educational Research; *School Surveys; Science Education; *Science Teachers; Secondary Education; *Secondary School Science; *Student Attitudes

Expanded abstract and analysis prepared especially for I.S.E. by Hans O. Andersen, Indiana University.

Purpose

The investigators' purpose was to study the relationships between students' perceptions of the ideal teacher and the student's prior success in science classes, gender of the student and the specific science course - grade level. The questions investigated were: Do science students rank teacher behaviors differently in their description of an ideal teacher according to a) success in science, b) gender of student and c) specific science course and grade level.

Rationale

When a teacher is perceived to be ideal by a student, it is assumed that the student learning will be optimized. Conversely a negative perception should have an opposite effect. That is why studies leading to understandings of how science students perceive ideal teacher behavior are of interest to classroom teachers, science education researchers and to teacher educators.

Research Design and Procedure

The study sample was 217 eighth (general science), tenth (biology), and twelfth (physics) grade students from the San Francisco Bay Area and from Salt Lake City, Utah. The students represented a range of SES-type schools. The Peterson-Yaakobi Q-sort which is composed of 24 single line descriptions of teacher behavior was the instrument used in this study. The Q-sorts were administered at the beginning of the second semester of their science class by one of the researchers. The data collected for the study included: a) sorts, b) gender of the student, and c) previous semester grade and course grade level. The independent variable was the mean ranks of ranks for each item by group. The other data constituted the dependent variables.

Findings

1. High Achievers saw four behaviors as more important for the ideal science teacher than did the low achievers. These were:
   #12 Acts like students are important as individuals
   #14 Quickly returns student work with comments or grades
   #20 Encourages and responds to students' opinions and ideas
   #24 Effectively gets across subject matter
2. Low achievers saw three other behaviors as more important for an Ideal Teacher than did the high achievers. These included:
   #2 Uses punishment to maintain control
   #3 Is disorganized
   #16 Uses words students can't understand
3. Males saw items #4 and #7 as more important behaviors of the ideal teacher than did females. These are:
   #4 Creates comfortable learning environment
   #7 Uses only test scores for grading
4. Females saw items #12 and #23 as more important behaviors of the ideal teachers than did males. These are:
   #12 Acts like students are important as individuals
   #23 Initiates contact with parents and community members
5. General Science, Biology, and Physics students perceived the Ideal Teacher behavior as significantly different on seven items, including:
   #2 Uses punishment to maintain control (General Science)*
   #4 Creates a comfortable learning atmosphere (Physics)*
   #6 Gives questions in tests which require memorizing (Biology)*
   #16 Uses words students can't understand (Biology)*
   #20 Encourages and responds to students' opinions and ideas (Physics)*

#21 Follows school rules and procedures (Biology)*

#24 Effectively gets across subject matter (Physics)*

*Ranked most important by students in stated discipline.

Interpretation

The authors drew the following conclusions.

1. Students differ significantly in their perception of an Ideal Science Teacher.
2. Perception is significantly related to class success, gender, and science subject under study.
3. Higher achievers ranked behaviors reflecting concern for learning higher than did low achievers.
4. Low achievers ranked statements of teacher behavior with negative connotations as well as relevance to learning higher than did higher achievers.
5. Differences in item ranking by gender groups can be said to follow stereotypic gender role expectations.

ABSTRACTOR'S ANALYSIS

In this study the investigators used a 24 item list of teacher behaviors and asked students to identify by ranking those behaviors they would associate with an Ideal Teacher. A twenty-four item list, and I am sure the authors would agree, is probably far too short to be an all inclusive list of teacher behaviors that would influence student perceptions. Nonetheless, the authors did uncover some very interesting relationships.

For example, what types of experiences have led the general science students to the perception that an Ideal Teacher uses punishment to maintain control? Is it because they prefer to be controlled rather than responsible for their own control? Or, do they not believe they can possibly control themselves? Of similar concern is the perception held by biology students that the Ideal Teacher uses memory questions on tests. For years, I have been concerned about biology teachers who make biology a foreign language class by emphasizing vocabulary and memorizing definitions. Is biology taught as a foreign language because of the teacher or because that is the way the students want to be taught? Who is shaping whom?

Also interesting is the fact that low achievers identified using punishment, disorganization, and pedantic language use as ideal behaviors of teachers. This seems to suggest that the low achiever has been convinced that school is good, experienced that school was personally painful, and then assumed that if school is not painful, it cannot be ideal. Need this perception be changed?

The data provided by the authors are most interesting. We need to know more about what students perceive to be Ideal Teachers. How the data will be used is important. Sometimes it will help guide our thinking about how teachers should behave. Sometimes it will help us identify student perceptions that need to be changed. Both actions are needed!

Chewprecha, T., M. Gardner and N. Sapianchai. "Comparison of Training Methods in Modifying Questioning and Wait Time Behaviors of Thai High School Chemistry Teachers." Journal of Research in Science Teaching, 17 (3): 191-200, 1980.

Descriptors--Autoinstructional Aids; Educational Research; *Inservice Education; *Independent Study; *Questioning Techniques; Science Education; *Science Teachers; Secondary Education; Secondary School Science; *Teacher Behavior

Expanded abstract and analysis prepared especially for I.S.E. by F. Gerald Dillashaw, Bradley University.

Purpose

This article reports a study conducted in Thailand to test the effectiveness of using audiotape models and instructional pamphlets as "self-training" materials to modify questioning and wait time behavior of Thai high school chemistry teachers.

Rationale

New science curricula emphasizing the role of the laboratory and inquiry learning were introduced into Thai senior high schools in 1976. The authors recognized that the new curricula placed more demands on teachers for new instructional strategies as compared to the traditional curricula. The authors argue that the use of high level questioning techniques is important to accomplish the goals of the new curricula. They indicate that Thai teachers primarily tend to ask factual recall type questions, as is typical of other teachers. The authors note that, in Thailand, it is difficult and prohibitively expensive to arrange inservice programs to assist teachers in developing new skills. Thus the study compared three different training models of self-instruction with Thai senior high school chemistry teachers.

Research Design and Procedures

The independent variable was the training method:

Method I-Study from Instructional Pamphlets. Teachers in this group were mailed three pamphlets to study -- one per month for three months. These pamphlets described purpose and value of questioning and suggestions on how to develop good questions.

Method II-Qualitative Listening to Audiotape Models. Teachers in this group received by mail three audiotapes of lessons with directions to "listen" to and "comment" on types of questions used in lesson.

Method III-Quantitative Listening to Audiotapes. This group received the same three tapes as Method II teachers, but rated the questions used in the model lesson according to a category system with which they had been trained.

There were five dependent variables measured in this study: (1) proportion of managerial questions, (2) proportion of rhetorical questions, (3) proportion of closed questions, (4) proportion of open questions, and (5) wait time (in seconds) following open questions. The questions were defined according to Blosser (1972) and wait time was defined according to Rowe (1974).

A pretest-posttest control group design was used in this study. Initially 102 chemistry teachers volunteered to participate in the study. However, only 77 teachers completed all tasks and these 77 comprised the final sample for data collection and analysis. Schools were randomly assigned to one of four groups: the three experimental groups and a fourth group which served as a control group. At the beginning of the study, all teachers provided an audiotape of the same chemistry lesson (Energy of Solutions) for the researchers. The dependent variables were assessed from these initial audiotapes by trained raters; these ratings served as the pretest measures.

Near the beginning of the first semester, 1976, teachers from the three experimental groups were invited to a central location for a two hour session on the importance of questioning and wait time. In addition, the Method III teachers were taught to classify questions in a further three hour session. During the first semester, Method I teachers were mailed three different instructional pamphlets (one per month) for study and response. Method II teachers were mailed three different audiotapes (also monthly) for listening and comments. Method III teachers received the same audiotapes with directions to classify and tally questions. The fourth group was the control group and received no treatment. At the beginning of the second semester, all teachers audiotaped the same lesson (Flame Test) in their regular classrooms. These tapes were analyzed by the researchers for frequency and type of questions and wait time; this analysis comprised the criterion measures.

Interrate agreement for the criterion measures averaged 95.8%. The proportion of types of questions was calculated and transformed using the Arc sine transformation. MANCOVA, using pretest scores as the covariate, was used to test for difference among the overall means of the four types of teacher questions. Following a significant MANCOVA, univariate ANCOVA was used to test for differences among the means for each type of teacher question. Following a significant ANCOVA, Newman-Keuls procedures were used for tests of pairwise contrasts. Wait time was analyzed separately using Newman-Keuls procesures again for follow-up.

Findings

A significant result for MANCOVA was found with the following findings reported for questioning behavior:

1) No significant differences among the groups in proportion of managerial and closed questions asked by the four groups of teachers.

2) Significant differences among the adjusted means of rhetorical and open questions asked. Teachers treated with Methods I and II asked significantly more questions than did control group teachers, but did not differ between themselves. Method III appeared not to have made a significant difference when compared to any of the groups.

3) Method I treatment teachers showed significantly more wait time compared to all the other groups.

Interpretations

The authors conclude that Method I (studying instructional pamphlets) and Method II (qualitative listening to audiotapes) were most effective in increasing the proportion of open questions in the classrooms with no real difference in effectiveness between the two methods. Method I was effective only in increasing wait time. The researchers had predicted that Method III (quantitative analysis of audiotapes) would be most effective; this prediction was not supported by the data. The researchers conclude that the demands placed on the Method III teachers were perhaps too extensive and skills in classifying questions not sufficiently developed. Hence, confusion about what actually was a particular kind of question may have been prevalent with Method III teachers.

Since Method I (instructional pamphlets) appeared to be most effective for improving both questioning behavior and wait time, the authors conclude that, for developing countries like Thailand, this approach may be very feasible, especially considering the low cost and ease of administration of such an approach. The authors speculate that Thai teachers eagerly used these written materials due to limited access to books and articles on education in Thailand.

ABSTRACTOR'S ANALYSIS

In general, the authors' exploration of the efficacy of various inservice training models for use in developing countries is a commendable one. The lack of extensive funds available for more elaborate inservice programs in such situations would seem to demand that educators in these areas try to find more cost efficient and effective means to upgrade skills of teachers.

There seem to be several potential problems with the study that call into question the conclusions of the authors. The use of volunteers as subjects represents a real threat to the generalizability of the results. The authors provide no evidence to indicate that this volunteer group was indeed representative of Thai chemistry teachers. Possibly they represent teachers who are most eager to improve their teaching skills. Additionally, the sample decreased from 102 to 77 during the study. The authors provide no explanation of the nature of the teachers who dropped out of the study. One must question whether this attrition was random or represented a particular group of teachers.

There are no problems with the data analysis. Procedures and interpretations of the statistical tests are sound. Method III may have the most serious problems in terms of the actual treatment. The authors state that the teachers appeared to have problems classifying the types of questions when compared to the raters' analysis of the tapes. If this be the case, then possibly this confusion contributed to the finding that Method III teachers did not improve questioning behavior. It is not clear to the reader what Method II teachers were to do when they were asked "to comment on the questions."

Readers familiar with wait time research should be aware that wait time, as used in this study, uses only Wait time I (pause after teacher talk). Rowe (1978) has also described Wait time II-pause after student talk. The teachers received only a discussion of the importance of wait time in that initial two hour orientation. There was apparently no systematic training to increase teacher wait time. This is a likely explanation for little increase in wait time behavior on the part of most of the teachers.

This study has attempted to find useful and efficient means of inservice training for developing countries such as Thailand where funds are scarce and logistical problems exist in delivering direct inservice training. The conclusions of the authors should be treated as tentative, due to some problems in the study as noted. It would seem that such a study, with the methodological problems minimized, would have great potential for addressing alternative inservice training models for use in such situations.

REFERENCES

Blosser, P. E. "The Development of the Skill of Questioning in Prospective and Secondary School Science Teachers: An Extension." Washington, D. C.: National Center for Education Research and Development, 1972.

Rowe, M. B. "Wait-time and Reward as Instructional Variables, Their Influence on Language, Logic, and Fate Control: Part One--Wait Time." Journal of Research in Science Teaching, 6, 49-53, 1969.

Riley, Joseph P., II. "A Comparison of Three Methods of Improving Preservice Science Teachers' Questioning Knowledge and Attitude Toward Questioning." Journal of Research in Science Teaching, 17 (5): 419-424, 1980.

Descriptors--Elementary School Science; *Elementary School Teachers; *Preservice Teacher Education; *Questioning Techniques; *Science Education; *Teacher Attitudes; Teacher Behavior

Expanded abstract and analysis prepared especially for I.S.E. by Thomas P. Evans, Oregon State University.

Purpose

The purpose of this study was to compare the effects of training in the use of three question classification systems on the ability of preservice teachers to classify written questions and on their attitude toward questioning.

Rationale

Inherent in the study was the assumption that it is desirable for teachers to learn and use a question classification system. The outcomes of such an activity include teachers having a better understanding of questioning and increases in the cognitive level of teacher questions and student responses. The assumption and outcomes were supported by a general statement about the findings of previous research investigations, making specific reference to Farley and Clegg (1969), Konetski (1970), Rogers and Davis (1971), Gallagher and Aschner (1963) and Riley (1978). The assumption, combined with the fact that a variety of question classification systems is available without specific data to support one system over the other, served as the contextual framework within which the study was conducted.

Research Design and Procedure

Twenty-seven female, preservice teachers in an elementary science methods course were randomly assigned to one of three treatment groups. Each group was provided training in classifying written science questions into one of three question category systems based on operationally defined hierarchies. The training session was composed of one, two-hour self-instructional module consisting of a description of each category followed by practice exercises in categorizing sets of science questions using the assigned hierarchy. A self-check was provided for each exercise. The treatment differed only in the module used for training, i.e., the Aschner, Bloom and Sanders Modules. The Aschner Module was based on Chapter Two of Blosser's Handbook of Effective Questioning Techniques (1973). The question hierarchy included the following categories: (1) Cognitive Memory, (2) Convergent, (3) Divergent and (4) Evaluative. The Bloom Module was based on Section One of Question Asking Skills for Teachers by Okey, Humphreys and Bedwell (1973). The question categories were as follows: (1) Knowledge, (2) Comprehension, (3) Application, (4) Analysis, (5) Synthesis and (6) Evaluation. The Sanders Module was based on Chapter Two of Hunkins' Questioning Strategies and Techniques (1972). It included the following question hierarchy: (1) Memory, (2) Translation, (3) Interpretation, (4) Application, (5) Analysis, (6) Synthesis and (7) Evaluation. The role of the instructor of the methods course was a passive one, responding only to clarify the modules' directions.

The treatment groups were pre- and posttested using the Science Question Classification Test to measure their ability to correctly classify questions into assigned category systems. The test consisted of a brief description of the cognitive levels and 40 science questions adapted from Appendix B of the Handbook of Effective Questioning (Blosser, 1973). The questions were to be categorized into the proper categories (Riley, 1978). The Attitude Toward Questioning Measure was administered to the treatment groups as a posttest to determine their attitude toward questioning. Both instruments had been developed by the investigator and judged to have face validity by a panel of experts.

Analysis

Scoring the Science Question Classification Test was different for each treatment group because the category system used by each group was not the same. As a result, measures of the dependent variable were expressed as percentages of correct responses on the posttest. The raw data were corrected for chance and transformed utilizing Arc sine transformations. Group means were analyzed using an analysis of covariance with the pretest as the covariate. The results of the analysis of covariance were subjected to the Newman-Keuls Multiple Comparison Procedure. Once the data were analyzed, they were retransformed into the original scale for interpretive purposes.

The scores of 22 subjects on the Attitude Toward Questioning Measure attitude scale were analyzed by means of an analysis of variance and the Newman-Keuls Multiple Comparison Procedure. The analysis did not include all the subjects because of missing data.

Findings

The findings reported by the investigator were as follows:

1. The pretest means, expressed as percent correct, on the Science Question Classification tests were 37.55, 31.60, and 23.33, respectively, for those groups who use the Aschner, Bloom and Sanders Modules;

2. The posttest means, expressed as percent correct, on the Science Question Classification Test, were 64.22, 39.70, and 33.44, respectively, for those groups who used the Aschner, Bloom and Sanders Modules;

3. The adjusted posttest means on the Science Question Classification Test for the three treatment groups were significantly different (P=.002).

4. The multiple comparison of posttest means, expressed as percent correct, on the Science Question Classification Test showed that subjects who used the Aschner Module score significantly

higher ($p<.05$) than those who used the Bloom or Sanders Modules.

5. A significant difference ($p=.08$) was reported among groups on an attitude subscale, which measured attitude toward the category system used by each treatment group, of the Attitude Toward Questioning Measure.

6. The multiple comparison of the treatment means on the Attitude Toward Classification System subscale of the Attitude Toward Questioning Measure show that subjects who used the Aschner Module ($\bar{X}$ = 24.05) were significantly different ($p<.05$) from those who used the Bloom ($\bar{X}$ = 29.37) or Sanders ($\bar{X}$ = 29.63) Modules.

Interpretations

The investigators concluded that the Aschner category system appears to be more effective than the Bloom or Sanders systems for selected outcomes when the subjects are restricted to a short training period. The selected outcomes include an improved ability to recognize and classify written questions into cognitive hierarchies and the development of a more positive attitude toward the classification system.

ABSTRACTOR'S ANALYSIS

The investigation by Riley was a logical and needed extension of an existing body of research that involves the systematic observation of science teacher questioning behavior. Earlier investigators have made available a variety of low inference observation instruments for identifying and classifying verbal classroom behavior, including the questions science teachers ask during instruction. These instruments have been used to describe the kinds of questions teachers ask and as training devices to modify teacher questioning behavior. Descriptive comparison of selected instruments are readily available in the research literature; however, as pointed out by Riley, specific data are not available to support the selection of one of these systematic observation systems over the other on the basis of specific outcomes. Riley's investigation has made a contribution to the existing body of research on science teacher questioning behavior. The selected outcomes or dependent variables and/or the instruments used to measure these outcomes might be questioned, but not the basic idea of the investigation. He is to be commended for his efforts to determine which of the three most common question classification systems is more effective as a training device in terms of teachers' ability to classify written questions and attitude toward questioning.

A major reason that the dependent variables and/or how they were measured might be questioned relates primarily to the research report. It was too brief. Critical details are omitted, and their omission makes interpretation and replication of the investigation very difficult, if not impossible. For example, at the end of the training period, were the teachers supposed to know the category systems well enough to classify the questions teachers ask, or were they supposed to be able to categorize questions while looking at a list of categories and their definitions? The latter situation was apparently true, but it could not be determined inconclusively from the report. These two possible outcomes of the training are quite different, and depending on the situation, would result in differences in the interpretation and replication of the investigation.

A second example of the research report being too brief relates directly to the Attitude Toward Questioning Measure. The instrument and how it was scored are not sufficiently described in the report. It could be inferred from a table and the discussion of the findings that the attitude measure was made up of individual subscales and that a low score on at least one of the subscales meant a more positive attitude. However, the number and nature of the subscales, including the number and nature of the items in each subscale, were unavailable. If the attitude measure has been developed specifically for this investigation, a sizable portion of the report should have been devoted to the instrument. If the measure had been used in earlier investigations or described in other documents, the report should have included references to those sources.

The use of the term "face validity" is a third example in which more information is needed in the research report. Riley provides identical statements about the validity of the criterion instruments; i.e., they were judged by a panel of experts to have face validity. The report contains a reference to an earlier investigation in which the Science Question Classification Test had been used, but the document did not provide any additional insight into the instrument's validity. Simply stating that the criterion instruments have face validity is not a sufficient descriptor of validity, because the term does not have a generally accepted meaning. Some researchers and specialists in educational and psychological measurement see validity as a misnomer. They feel it is not a type of validity at all. Some see it as a superficial technique involving a visual inspection and description of whether or not an instrument appears to measure what it is supposed to measure. Some feel that, although face validity is not a rigorous concept, it has value when used in conjunction with and reinforces content, criterion-related or construct validity. Others confuse it with or use it as a synonym for content validity. Was face validity used in the investigation as a synonym for content validity or in conjunction with one of the other types of validity? If not, the validity of the criterion instruments is suspect. In either situation, interpretation of the investigation cannot be properly accomplished without additional information concerning the validation of the criterion instruments.

It should be pointed out that face validity is not only a vague but an outdated term. In 1966, the American Psychological Association (French and Michael, 1966) recommended that only three types of validity be used in educational and psychological measurement, namely, content, criterion-related and construct validity. The intent of the recommendation was to bring about a set of common language and expectations concerning validity. Researchers and those who are involved with measurement in science education would do well to follow this recommendation.

A fourth example where more information was needed to properly interpret and replicate the investigation related to the modules. More details describing the modules are necessary before their differences, beyond the question category systems, can be ruled out as an alternate explanation for the significant differences that were found among the treatment groups on the dependent variables.

Charging that the research report was too brief is not meant merely as a criticism of the author, or the journal and its editor. The problem seems to be universal. Journals have limited space, and decisions must be made about what is to be included or deleted from a report. Frequently, details are omitted which are necessary for interpretation because of the author's familiarity with the research. At other times, deletions represent the lack of an editor's understanding of the investigation. Regardless of the reasons, researchers in science education and editors should give more thought to what is included and deleted from a report. In Riley's case, the report would have been improved if several of the tables had been omitted and the space devoted to further descriptions of the modules and criterion instruments. Another possibility would have been to submit two cross-referenced reports, each emphasizing a different aspect of the research in detail.

The investigation by Riley brings to mind a number of ideas for future research involving the systematic observation of science teacher questioning behavior. There is some evidence to support the position that the ability to classify written questions is positively related to the level of questions science teachers ask in the classroom, but the evidence is not conclusive. Therefore, the most obvious investigation that needs to be conducted is to compare the effectiveness of the individual category systems as training devices for modifying the actual classroom questioning behavior of science teachers. The effectiveness of the question category systems could also be compared using other selected teacher and student outcomes as well. In looking at these outcomes, some of the limitations of the process-product research strategy might be overcome by employing a research strategy similar to the one posed by Medley (1982) for investigating teacher effectiveness. The new strategy should include a consideration of other variables that might influence the outcomes such as teacher, group and school characteristics. The ability to classify questions also needs further investigation. Learning to categorize written questions with the aid of a list of categories and definitions may influence outcomes differently than actually learning to identify and categorize teacher questions, live or on tape, with a high level of inter-observer agreement.

REFERENCES

Blosser, Patricia E. Handbook of Effective Questioning Techniques. Worthington, Ohio: Educational Associates, Inc., 1973.

Farley, G. T. and A. A. Clegg. "Increasing the Cognitive Level of Classroom Questions in Social Studies: An Application of Bloom's Taxonomy." Paper presented at the American Educational Research Association Symposium, "Research in Social Studies Education." Los Angeles, February, 1969.

French, J. W. and W. B. Michael (Co-Chairmen). Standards for Educational and Psychological Test and Manuals. Washington, D.C.: Psychological Association, 1966.

Gallagher, J. J. and M. J. Aschner. "A Preliminary Report: Analysis of Classroom Interaction." Merril-Palmer Quarterly, 9: 183-194, 1963.

Hunkins, F. P. Questioning Strategies and Techniques. Boston: Allyn and Bacon, 1972.

Konetski, L. C. "Instructional Effect on Questions Asked by Pre-Service Science Teachers." Dissertation Abstracts, 30: 4858A, 1970.

Medley, D. M. "Teacher Effectiveness." Encyclopedia of Educational Research. Edited by Harold E. Mitzel. New York: The Free Press, 4: 1894-1903, 1982.

Okey, J. R., D. W. Humphreys and L. C. Bedwell. Question Asking Skills for Teachers. Bloomington, Indiana: National Center for the Improvement of Educational Systems, Indiana University, 1973.

Riley, Joseph P., II. "Effects of Studying a Question Classification System on the Cognitive Level of Preservice Teachers' Questions." Science Education, 62: 333-338, 1978.

Rogers, W. M. and O. L. David. "Varying the Cognitive Levels of Classroom Questions: An Analysis of Student Teachers' Questions and Pupil Achievement in Elementary School Social Studies." Paper Presented at the meeting of the American Educational Research Association, Lexington, 1971.

Sherwood, Robert D. and Dorothy Gabel. "Basic Science Skills for Prospective Elementary Teachers: Measuring and Predicting Success." Science Education, 64 (2): 195-201, 1980.

Descriptors--Anxiety; College Science; *Course Evaluation; Elementary School Science; *Elementary School Teachers; Higher Education; *Preservice Teacher Education; *Science Curriculum; Science Education; Science Teachers; *Teacher Attitudes

Expanded abstract and analysis prepared especially for I.S.E. by Leon Ukens, Towson State University.

Purpose

The purpose of this study was to determine the effectiveness and to predict success in a course for preservice elementary teachers at Indiana University entitled "Basic Science Skills." The course has three major components: science process skills, mathematical skills, and a unit on the structure of matter. "Basic Science Skills" is a prerequisite to four other required courses. The unstated hypothesis tested was one of determining course effectiveness, especially in the area of attitudes.

Rationale

Courses for preservice elementary teachers have been developed based upon a number of different models. This course was designed to have an impact in the areas of attitudes toward science and science teaching and, because of a concern for the mathematics needed for success, attitudes toward math. It was a course developed with the rationale of getting the students off to a good start for successful completion of subsequent courses by emphasizing science process skills.

Research Design and Procedure

The sample consisted of 105 students at Indiana University of whom 80% were freshman or sophomores and 92% were female. The study was carried out within a two-semester period. The "Basic Science Skills" course consisted of one 45-minute lecture per week and two two-hour lab sessions. The content was selected by the scientists teaching the course and the science education faculty.

Units included observation, inference and prediction, variables, graphing, measurement and the metric system, large and small numbers, proportion, problem solving, hypotheses, operational definitions, and the nature of matter. Activities were drawn from SAPA, SCIS, ESS, IPS, and other textbooks.

Six instruments were administered to the students. These were (1) Demographic Data Questionnaire, (2) Fraction and Decimal Test with KR-20 of 0.70, and (3) a Math Anxiety Test with a reliability of 0.97, administered as pretests, (4) a Basic Science Skills Test, and (5) Attitude Toward Science and Science Teaching Tests, with reliabilities of 0.93 and 0.83, respectively, administered as pre- and posttests, and (6) three Content Achievement Tests, with reliabilities of 0.58 and 0.71, respectively, administered as posttests.

Three different analyses were performed to evaluate course effectiveness and to predict success. These were the correlated t-tests of the pre- and posttest scores of the Basic Science Skills Test (both semesters), a correlated t-test of the pre- and posttest scores on the Attitude Toward Science and Science Teaching tests (first semester only), and a multiple linear regression analysis producing an equation for predicting success.

Findings

There were significant differences in pre- and postttest scores on the Basic Science Skills Test and no significant differences between attitudes as measured by the Attitude Toward Science and Science Teaching test.

In developing the regression equations, two measures were used as dependent variables: one was total course points, a combination of quiz scores, lab report grades, a lab practical, and tests, and the other was total test points. For total test points the Fraction and Decimal test was the best predictor of success, followed by number of math courses in high school, the Math Anxiety Test and the Basic Science Skills pretest. These four accounted for 43.0 percent of the variance. For total course points the first two were the same, but the Math Anxiety Test and Basic Science Skills test were replaced by the negative attitude toward science subscale from the Attitudes Toward Science and Science Teaching test.

Interpretations

The "Basic Science Skills" course's objective of improving attitudes toward science and science teaching was not realized. This could have been caused by a number of factors: (1) the Attitudes Toward Science and Science Teaching test may not have been sensitive enough to detect changes, or (2) the content of the course was arranged with the mathematics part being the last third of the course directly prior to the time that the instrument was given as a posttest.

The prediction of success in the course was heavily influenced by mathematics variables. This may mean the course reflected a level of precision and thinking not encountered in the students' previous courses in high school and college.

Even though the Basic Science Skills test showed a learning of these skills, the authors were still disappointed with the results.

ABSTRACTOR'S ANALYSIS

An important concern for educators is determining the various ways a course may be effective. The authors of this study were interested in the important attributes of attitudes toward science and science teaching among other outcome variables. Their finding of no significant attitude change as a result of the "Basic Science Skills" course needs closer examination. As the authors pointed out, the reason may be that the changing of attitudes over such a short period of time may not be possible, or it may be that the Moore attitude instrument may not be sensitive enough to measure such changes. Another reason, not mentioned, is the nature of the students in the course. By the time a student gets to be a freshman or sophomore in college, attitudes toward school and school subjects are already deeply set, and changing these over such a short period of time may be impossible or superficial at best. But suppose for a moment that the attitudes did change, could that change be attributed to this particular course? Or could the change have been due to some other variable? Controlling significant events in the lives of college students while trying to determine the effectiveness of one course is virtually impossible to do. Therefore, it seems that attitude changes among college students as a result of any course would be suspect. It would be interesting to test a group of students in a more traditional lecture course and determine if and how their attitudes change.

Readers of this study get some insight into the "Basic Science Skills" course described but not enough to determine how the course is similar to or different from courses in their own schools. To report the details of such a course would take numerous pages in the journals, however, if readers of these evaluations are to do more with their own curruculum development endeavors, more information concerning the specifics of the courses are needed. This criticism is not just of this study, but of nearly all others reported in the same fashion. About all the reader can obtain is a general course description along with a model for evaluation.

This reviewer has used the Moore test on Attitudes Toward Science and Science Teaching as a pre-, posttest in evaluating a physical science course developed for preservice elementary teachers, and the results were just about the same as reported here. It probably does depend, as the authors pointed out, on what was done in the course prior to the posttest. Perhaps another on-going, less detailed instrument could be used.

The idea of predicting success in a course for the purpose of determining help needed is a sound one. However, the way this study is reported, it almost seems as an afterthought. No mention was made on how the predictors were used except to develop a regression equation.

In conclusion, while studies of course effectiveness are worthwhile for particular courses, students, and instructors, the results are not generalizable much beyond this. This study is no exception.

Cotham, Joseph and Edward L. Smith, "Development and Validation of the Conception of Scientific Theories Test," Journal of Research in Science Teaching, 18 (5): 387-96, Sept. 1981.

Descriptors--Elementary Secondary Education; Evaluation; Science Education; Science Teachers; *Science Tests; Scientific Concepts; *Test Construction; *Test Validity; *Theories

Expanded abstract and analysis prepared especially for I.S.E. by Rodney L. Doran and Edward S. Jenkins, SUNY at Buffalo.

Purpose

The authors sought to "develop a reliable and valid instrument for use with elementary and secondary teachers of science" that would assess their conception of the meaning of science. They set out to do this by constructing an instrument that would be (1) "sensitive to alternative conceptions of particular philosophic aspects of scientific theory and (2) could be "[Useful] in inferring understanding of the tentative and revisionary conception of the nature of science".

Rationale

The authors contend that the teachers' conception of science has educational and social importance. Referenced to the educational importance, they aver that the teachers' conception of science significantly influences "teaching behavior." They cite as example that the teaching style of an inquiry-oriented teacher, for instance, would reflect that view. Moreover, the curriculum designed to present science to students as inquiry is best accomplished by the teacher who considers and understands science to be a discipline of inquiry. On the other hand, if science is presented as a "collection of immutable facts," rather than emphasizing the "inconclusiveness of all knowledge claims," then there is bound to be an erosion of confidence in, and loss of support by, society when new claims are made by "such rapidly developing disciplines as astrophysics, nuclear physics and biochemistry." Thus, the social implications.

Cotham and Smith assert that teachers' conception of science can be measured, but instruments extant are inadequate for the task. They advance the notion that the needed investigation of the "relation between teachers' conception of the nature of science and how they instruct science" can be realized by an instrument that is (1) "sensitive to alternative conceptions of the particular aspects of the nature of science" and (2) useful in "inferring possession of a tentative and revisionary conception of the nature of science."

Research and Design

Referenced to many existing instruments for measuring teachers' conception of science, the authors cited their "single interpretation" as a failing. Teachers, they believe, are also likely to embrace several false conceptions and, to these, any measuring instrument must be sensitive. Therefore the development of the Conceptions of Science Theories Test (COST) was aimed at meeting two critera: (1) "sensitivity to alternative conceptions of the aspects of the nature of science" and (2) "sensitivity to a conception of science that has particular relevance to teachers of science." They developed an "attitude" instrument consisting of modified Likert items and four subscales: (1) ontological implications of theories; (2) testing of theories; (3) generation of theories; (4) theory choice. The reference point for each subscale was a philosophical issue with two alternative conceptions for each issue. The response scale [(1) strongly agree, (2) agree, (3) disagree, (4) strongly disagree], and test items were arranged so that agree responses were consistent "with 'conclusive' alternatives and disagree..." with 'tentative' alternatives. Contextual items included: (1) Bohr's theory of the atom; (2) Oparin's theory of abiogenesis; (3) Darwin's theory of evolution; (4) theory of plate tectonics; and (5) non-theoretical context. A brief description of a scientific theory and some "episode from its history" were followed by test items. The constructs were largely based on the philosophies of Hempel, Kuhn, Martin and Nagel. They adopted Meehl and Cronbach's definition of construct: "some postulated attribute of people assumed to be reflected in test performance." Two pilot forms, A and B, each consisting of 40 items, were administered to 56 physical science students (form A to 29 and form B to 27). Following analysis of results, they then selected 40 items judged to exhibit the strongest subscale traits. The items constituted the final form of COST. Three populations of students, based on identified majors, were selected to take the test so as to ascertain the discrimination capacity of the instrument. Next, the multi-method procedure of Campbell and Fiske was used for examination of subscale validity.

Based on assumptions pertaining to differences between elementary education students and philosophy students, on the one hand, and elementary education students and chemistry students, on the other, in terms of their relative "sophisticated conception of theory testing and theory generation," five hypotheses were formulated:

1. Elementary education majors will perform according to a "conclusive" conception of the testing theories more than philosophy students.
2. Elementary education majors will perform according to an "inductive" conception of the generation more than philosophy students.
3. Elementary education majors will perform according to an "objective" conception of theory choice more than philosophy students.
4. Elementary education majors will perform according to a "conclusive" conception of testing of theories more than chemistry students.
5. Elementary education majors will perform according to a "inductive' conception of the generation of theories more than chemistry students.

Tests were administered to: (1) 50 elementary education majors; (2) 30 chemistry and chemical engineering majors and (3) 30 (volunteer) philosophy of science students in a major midwestern university. Construct validity determination consisted of the following: First, a one-tailed test of significance was used to determine whether or not there was a significant difference between subscale scores of the elementary education majors and the other groups. Secondly, the Campbell and Fiske multitrait-multimethod matrix procedure was employed to determine and present correlation measures between subscales.

Findings

1. The t-values obtained indicated a significant difference "in the predicted direction" between elementary education majors and philosophy majors ($p < 0.01$).
2. Similarly t-values obtained indicated signficant difference "in the predicted direction," between elementary education majors and chemistry majors ($p < 0.01$).

Referenced to construct validity, the authors present a signficant validity coefficient mean score of 75% (range = 60-100), significantly greater than zero. The mean scores of smaller same method coefficient and smaller different method and trait coefficients were 43% and 45% respectively, each significantly smaller than the validity coefficient at 0.10 level (ranges 20-50% and 29-47% respectively). Values obtained revealed highest scores for the testing of theories subscale and lowest for theory choice. Reliability estimates were not published because, according to the authors, there was a considerable range of variance for the three groups. They only stated that the standard error was "relatively" lower, suggesting that this inferred a degree of reliability interpretations.

The test results published by the authors supported their hypotheses that elementary education majors differ from chemistry majors and philosophy of science majors in their conception of science, viz.,

(1) testing of theories and (2) generation of theories. Elementary education majors differ from philosophy of science majors in theory of choice. The authors also asserted that the construct validity of COST supports the claim that their instrument measures selected aspects of scientific theories. Validity coefficient for the theory choice subtest was rather low, but, they aver, the absence of attention to this aspect of science in other instruments, an awareness of considerable heterogeneity in responses recorded by those taking the test, along with what the authors see as "the importance of assessing teachers' understanding of the tentative and revisionary conception," argues convincingly for the utility of COST. They regard their instrument as being appropriate for providing useful information on teaching effectiveness in relation to their concept of science. While they caution that interpretation of test results must recognize existing controversy pertaining to "particular aspects of the nature of science," they nevertheless believe that their instrument is useful in identifying factors pertinent to the teaching of science as inquiry - a pedagogical approach they obviously espouse.

ABSTRACTOR'S ANALYSIS

This study was based on the premise that teachers' conception of the nature of science shapes teaching style or practice. The authors contended that extant instruments designed to investigate teachers' concept of science lacked sensitivity by limiting a user's response to a single interpretation. They moved to correct what they regard as a deficiency by constructing an instrument which allowed responses to alternative theories, so as to signficantly improve the discriminating qualities of the test. Even so, there is a measure of limitation in the pre-selection of the alternatives included.

The philosophical views of Hempel, Kuhn, Martin, and Nagel were chosen as construct criteron sources. As was done for Kuhn, the authors might have provided a basis for their choices of philosophers of science.

The authors classified three different student populations based on disciplines. However, one might question whether or not chemistry and chemical engineering majors form a homogeneous group, and whether the fact that one group, philosophy of science majors, were the only volunteers introduced extraneous factors. Interpretation of the data pertaining to test discrimination must recognize this fact as well as the authors' acknowledged limited populations involved. A test to be used with teachers should be validated with teachers. Perhaps that is the second stage on their research effects. No validity or reliability coefficient scores are presented. However, the authors present their multi-method, multi-trait analysis in terms of the percentage and percentage mean subscale scores with validity coefficient "significantly greater than zero." This came to 100% for testing of theories, 80% for generation of theories and 60% each for autological interpretation of theories and theory choice. The mean percentage was 75%. However, the authors might have rendered some comparisons between the construct validity inferences of their test to others which they reviewed in preparation for the investigations they made. The discussion of the validity requirements 2 and 3 was scanty, both in terms of the expectations and the results. In presenting data on discriminant validity, smaller coefficient values were obtained for each subscale and the ratio of measures for the theory choice was, as acknowledged by the authors, "comparatively low." Perhaps this section of the test might be deleted or revised.

Neither reliability estimates nor standard errors were tested, possibly in the interest of space economy. However, some samples might have been included in the article. The authors averred that test reliability was acceptable because all standard errors were less than 0.3. For this test characteristic also, some comparison with other related tests might have been beneficially included.

Finally, the authors state that "investigations of factors that influence successful inquiry teaching should certainly assess teachers' understanding of the tentative and revisionary conception of science, a conception of science intimately related to understanding of 'science as inquiry'." However, in the face of their oft-repeated assertion that

their intent was to investigate "the relationship between teacher's conception of science and teaching effectiveness," the foregoing appears somewhat biased. An apparent typographical error somewhat obscured their statement of implications for the test in "inquiry teaching." In summary, let it be said that the authors have made a good start and a solid contribution in a most difficult to assess, but important to measure, aspect of science education.

Gabel, Dorothy. "Attitudes Toward Science and Science Teaching of Undergraduates According to Major and Number of Science Courses Taken and the Effect of Two Courses." School Science and Mathematics, 81 (1): 70-76, 1981.

Descriptors--*Academic Achievement; College Science; Educational Research; Higher Education; Majors (Student); Nonmajors; *Preservice Teacher Education; Scientific Attitudes; Science Curriculum; Science Education; *Science Instruction; *Student Attitudes; Undergraduate Students

Expanded abstract and analysis prepared especially for I.S.E. by Edmund A. Marek, The University of Oklahoma.

Purpose

The purposes of this study were to: 1) compare teacher-education majors with other majors with respect to attitudes toward science, attitudes toward science teaching and course achievement; 2) determine whether the number of courses studied influenced attitudes and whether students enrolled in a special course relating the science studied to science teaching influenced attitudes; 3) measure the effect of participations in a science course designed for nonscience majors on attitude change; and 4) provide a basis for changing an extant science program for elementary education majors.

Rationale

Research (Korth, 1969; Hirschhorn, 1974; and Tilford and Allen, 1974) has shown science majors to have more positive attitudes than nonscience majors and further research (Jackson, 1968; Gabel, Rubba and Franz, 1977; Johnson, Ryan and Schroeder, 1974; and Kennedy, 1973) has shown that special approaches for teaching science courses can result in positive attitude development.

The author documents the growing concern by teachers and the general public over the development of positive attitudes toward science and teaching science. This is important in courses for nonscience majors and especially important for students who will eventually teach science.

Research Design and Procedure

The research design and procedures of this investigation were conducted in four phases: 1) attitudes toward science and science teaching were analyzed according to the student's major and science background, 2) achievement in the course according to major and science background, 3) effect of a treatment (a particular geology course) on attitudes toward science and 4) effect of a treatment (a special course for elementary teachers) on achievement and attitude.

The geology course included an overview of both physical and historical geology with an emphasis on scientific method. The course consisted of two 45 minute lectures and one 1 3/4 hour laboratory per week for three semester credit hours. The specially-designed education course met two hours per week for one semester credit hour and was essentially an elementary science methods course with a geology emphasis.

Attitudes were pretested and posttested with Moore's "Science Teaching Attitude Scales." Data were analyzed using a one way ANOVA. (Survey questions, reliabilities and validity procedures were described in the article).

The subjects of this study were 189 students enrolled in an introductory geology course for nongeology majors. Twenty-four of these students were elementary education majors, and 12 of the 24 students elected to enroll in the specially designed education course.

Findings

For the purposes of this research the students were grouped into one of four categories of majors: elementary education (26 students), science-teaching (8 students), science-nonteaching (12 students) and others (143 students). There exist significant differences, with science majors intending to teach in secondary schools having the most positive attitudes toward science and teaching science. There exist no significant differences between elementary education majors of this study and other majors of this study on attitudes toward science or toward teaching science. The investigator reports significant differences between attitudes and science background, with students for whom this was a first or second science course having the lowest attitude score.

Teaching and nonteaching science majors had significantly (0.01 level) higher achievement in the science course than did elementary education majors and other majors. Achievement in this course was also related to science background, with students enrolled in this geology course as a first science course achieving significantly (0.01) lower than any other students in the course.

There exists no increase in attitude toward science due to the treatment (geology course). There also exists no significant effects of attitudes and achievement of elementary education majors due to the specially designed education course which related geology to teaching.

Interpretations

Interpretation and major conclusions of thus study advanced by the investigator are several. Science majors have more positive attitudes toward science than do nonmajors. Elementary education majors' attitudes are no different than those of nonmajors. The number of science courses taken has a positive effect on student attitudes. This finding suggests that students should be required to take at least four courses in science (according to this study) if positive attitude development is a program goal. The students of this study who took more science courses also demonstrated greater achievement.

The specially designed education course (applying geology to teaching) had no effect on student attitude and course achievement. The researcher is cautious in this interpretation because of the small sample size for this phase of the project and the nonrandom student grouping. The author concludes that a program for elementary majors should include at least four science courses and special attention should be devoted to students' attitudes toward science and science teaching.

ABSTRACTOR'S ANALYSIS

Several dimensions were designed into this research to examine attitudes toward science, attitudes toward science teaching, science background, academic major and the effect of a treatment -- a science course and a specially designed education course. Most of the results from this study -- regarding attitudes, academic majors and science course background -- were predictable.

Some results, however, were unexpected. The treatment (courses in geology and education) produced no significant effects on attitude and achievement according to this study. Since one of the reasons for undertaking this study, according to the author, was to provide data for program change and development, it is assumed that the treatment, the courses as currently designed, will not remain part of the teacher preparation program. The abstractor recommends that the investigator reexamine this issue and correct the deficiencies of this study: small sample size and lack of random assignment into groups.

Additional research may provide useful data for redesigning the study. Marek and Lewis (1980) conducted research to examine: 1) What attitudes exist among preservice elementary teachers toward science? 2) What attitudes exist among preservice elementary teachers toward their science content courses? 3) What is the academic and biographical background of the preservice elementary teachers of the study? and 4) Do relationships exist among attitudes toward science, attitudes toward the required science courses, grade point average and gender of the preservice elementary teacher? This research (Marek and Lewis, 1980) provides additional fundamental data on attitudes and preservice teachers. Furthermore, research procedural steps, instrumentation, and additional research questions are reported in the study which may provide useful alternatives to the investigator. These research questions are currently being analyzed: 1) Can treatment (activities in science laboratory classes) affect the attitudes of preservice elementary teachers?, 2) Do preservice elementary teachers intend to teach science?, and 3) Does a relationship exist between intentions to teach science and attitudes toward science?.

Yager (1978) identifies the importance of considering attitudes in: "Priorities for Research in Science Education: A Study Committee Report:"

> The self-image and aspirations of teachers, and their images of science, of children, of learning and of effective activity are suspected of being highly predictive of the kinds and effectiveness of transactions which occur in school and the subsequent outcome. Therefore, teacher characteristics and their influence upon the transactions and outcomes has the highest priority for research ... Independently, but simultaneously, the Research Committee of the National Association for Research in Science Teaching (NARST) assigned the highest priority to empirical tests of the relationships between teacher attitudes and behaviors and student outcomes (Yager, 1978).

REFERENCES

Gabel, D., P. Rubba and J. Franz. "The Effect of Early Teaching and Learning Experience on Physics Achievement, Attitude Toward Science, and Science Teaching and Process Skill Proficiency." Science Education, 61: 503-511, 1977.

Hirschhorn, J. "Engineering and Nonengineering Students' Attitudes Toward Science and Technology." Science Education, 57: 29-33, 1974.

Jackson, P. Life in Classrooms. New York: Holt, Rinehart and Winston, 1968.

Johnson, R., F. Ryan and H. Schroeder. "Inquiry and the Development of Positive Attitudes." Science Education, 58: 51-56, 1974.

Kennedy, T. "The Effect of Process Instruction on Changing Preservice Elementary Teachers' Attitudes Toward Science." School Science and Mathematics, 78: 569-574, 1983.

Korth, W. Understanding the Social Aspects of Science. ERC Papers in Science Education No. 1. Cleveland, OH: Educational Research Council, 1969.

Marek, E. A. and M. Lewis. "Attitudes of Preservice Elementary Teachers in General Science." Journal of Professional Studies, 5 (3): 9-18, 1980.

Tilford, M. and D. Allen. "Science and Nonscience Majors in Three Predominantly Black Colleges." Journal of Negro Education, 43: 117-126, 1974.

Yager, R. E. "Priorities for Research in Science Education: A Study Committee Report." Journal of Research in Science Teaching, 15: 99-107, 1978.

MISCELLANEOUS

ARTICLES

Fisher, K., S. Williams and J. Roth. "Qualitative and Quantitative Differences in Learning Associated with Multiple-Choice Testing." Journal of Research in Science Teaching, 18 (5): 449-464, 1981.

Descriptors--Academic Achievement; College Science; College Students; *Computer Assisted Testing; *Feedback; Genetics; Higher Education; *Learning; *Multiple Choice Tests; *Retention (Psychology); Science Education; Science Instruction; Science Tests; *Student Attitudes; Testing

Expanded abstract and analysis prepared especially for I.S.E. by William M. Frase, University of Cincinnati.

Purpose

The purpose of this study was to assess the effects of a frequent multiple-choice testing system with immediate computer feedback for learning in an upper division genetics course for undergraduates by means of a frequent individualized testing procedure. Further efforts were made to measure this system for learning retention as well as measuring "meaningful" vs. "rote" learning. The paradigm was comparison of lecture/auto-tutorial classes in similar classroom situations (lecture, testing, instructional aids, instructional support, etc.) where one group received Computer Assisted Self-Evaluation (CASE) as compared to mid-term/final evaluation. CASE students received immediate feedback on quiz testing while mid-term students experienced a delayed feedback.

The null hypotheses of this study were:

1) There is no difference in attrition (drop out) rates between the two sections.
2) There is no significant difference between groups in knowledge of genetics as measured by a pre-test administered on day one.
3) There is no significant difference between groups in overall learning of genetics as measured by mean scores on the final examination.
4) There is no significant difference between groups in overall retention of genetics knowledge as measured by mean scores on a two-year retention test.

5) There is no significant difference between groups in rote learning of genetics as measured by mean scores on fact/recall items on the final examination.
6) There is no significant difference between groups in retention of rote knowledge of genetics as measured by mean scores on fact/recall items on the retention test.
7) There is no significant difference between groups in meaningful learning of genetics as measured by mean scores on problem-solving items on the final examination.
8) There is no significant difference between groups in retention of meaningful knowledge of genetics as measured by mean scores on problem-solving items on the retention test.
9) There is no difference between groups in their attitudes towards the course.

Rationale

Inherent to the purpose of this study was an attempt by the researchers to add to the theoretical body of knowledge that learning can be enhanced by precise, frequent and immediate feedback to the learner as found in the works of Pressey (1950), Skinner (1954), and Ammons (1956). An additional attempt was made to prove the merits of frequent questioning and feedback to the learner as found in programmed/audio tutorial/personalized instruction systems of Skinner (1958), Postlethwait (1963) and Keller (1968) which included affective measurements of student preferences for this type of instruction as opposed to more traditional strategies. Finally, an effort was made to examine the quality of learning which occurred, focusing on the various cognitive levels as identified in the studies of Bloom (1956), Gagne (1970) and Ausubel (1968).

The model for this research was a classic comparison of a traditional control group with that of a one variable (testing) experimental group. Class size, amount of lecture time, class materials, pre-test, final exam, and similar mid-term/quiz questions were constants. Each section met weekly at the same hour; each received three units of credit; and the quiz group took 24 quizzes for 60% of their grade, whereas the mid-term group took two mid-term of 120 questions between the pre-test and final. The significance of differences between groups was determined by t-tests and values of less than 0.025 were considered significant.

Research Design and Procedures

The basic design was to take two groups of undergraduate upper division students enrolled in a genetics course and employ treatment differences in the methods used for testing and feedback. The control (hereinafter referred to as M) consisted of 37 students who took a traditional mid-term form of evaluation and the experimental group (hereinafter referred to as Q) that took the CASE method of testing and feedback. The course of instruction lasted for ten weeks and lectures were about topics selected for their motivational potential. Both sections received lectures on all but the fourth and eighth week when the M group received its two mid-terms while class meetings for the Q group were cancelled. The CASE System was used for individualized testing of Q group students, for generating mid-terms for the M group and for producing the final exam for both groups. Twenty-four topics were covered in the course, and the same test items were administered equally to both groups ranging in difficulty from 0.4 to 0.95 with most in the 0.6 to 0.8 range. Test question validity was inferred from the fact that a number of genetics faculty members had used the items over a period of several years and found them appropriate for the material covered in this course. Reliability of test items came from the stability of item statistics (difficulty, discrimination, matrix analysis) over successive quarters of use. Each Q group quiz consisted of four questions on the topic of study and one question on the previous topic. Test versions for each quiz were generated randomly by computer. Q group students took their quiz in a testing center at their convenience with immediate computer feedback following quiz completion. M group students took mid-terms during regularly scheduled class periods in the first, fourth, and eighth weeks of instruction and received feedback several days later. The post-test was the same for both Q and M groups and consisted of one item from each of the 24 course topics and 16 items from a review file for a total of 40 questions. The retention test consisted of 24 items selected from the post test with one from each of the 24 topics taught. It was administered under the honor system to 48 students from both groups. They were all who could be located two years following the completion of the aforementioned course.

All test items used were classified independently by five instructors of genetics for three cognitive levels:

1) Those that were clearly rote (fact-recall).
2) Those that were clearly meaningful (problem-solving).
3) Those that were intermediate.

Mean scores and standard deviations were calculated for both Q and M groups within each of the three cognitive categories on the post retention tests.

An independent attitude assessment was taken by means of a survey distributed during the last class meeting which solicited student opinion on course content, televised programs, testing, instructors, textbooks, and syllabus. Most items used a Likert scale with numerical values assigned from 1 (strongly disagree) to 5 (strongly agree). Mean responses were calculated as if the scale represented a continuum.

Findings

The treatment results of the Q and M groups as related to previously stated null hypotheses are as follows:

1) There is no difference in attrition rates between the two sections. This was not rejected as group Q & M had similar attrition rates of 21% and 19% respectively.
2) There is no significant difference between groups in knowledge of genetics as measured by mean scores earned on a 24 item pre-test administered on the first day - was not rejected. Mean group scores on the pre-test were equivalent for groups Q & M with scores of 57% and 60% respectively.
3) There is no significant difference between groups in overall learning of genetics as measured by mean scores on the final examination was rejected by virtue of an overall level of significance (t = 2.86) for performance on the 40 item post-test. The Q group earned a mean score of 75% as compared to 66% for the M group.
4) There is no significant difference between groups in overall retention as measured by mean scores on a two year retention test - was not rejected. It should be pointed out that although the two groups had similar levels of performance on the retention test, (79% for Q compared to 72% for the M group) a level of p less than 0.10 (t = 0.89) was obtained and is almost significant.
5) There is no significant difference between groups in rote learning as measured by mean scores on fact-recall items on the 40 item post-test.
6) There is no significant difference between groups in retention of rote knowledge as measured by fact-recall items on the retention test. This was not rejected - was rather relegated to either accepting or rejecting the null hypothesis. There was insufficient information to for a view.
7) There is no significant difference between groups and "meaningful" learning of genetics as measured by mean scores on problem solving items from the final examination. This was rejected. The magnitude of difference between the Q and M groups on those

items labeled as measures of meaningful learning was significant (t = 3.93) for the post-test as well as the unique subset (t = 2.42).

8) There is no significant difference between groups in retention of meaningful knowledge as measured by mean scores on nine items on the two year retention test. This was tentatively accepted.

9) There is no difference between groups in their attitudes towards the course. This was rejected. Student attitudes showed that the majority of students in both groups favored the CASE quizzing over traditional mid-term examinations. Questions regarding knowledge about genetics, overall course evaluation, motivation, text, use of testing for learning, method of testing, and frequency of feedback were all measured attitudinally.

The researchers concluded that students learn significantly more with the CASE Quizzing System of 24 quizzes per quarter with immediate feedback than from traditional approach of two fifty minute examinations per quarter with delayed feedback for an overall quantitative gain in learning.

Findings of this study about the cognitive level of learning promoted by multiple choice tests are contrary to common belief. This may be largely due to a confounding of question type and question level. The tentative conclusion is that when students are given frequent multiple choice tests with items across the span of cognitive levels typically employed in an introductory science course, meaningful learning is enhanced at least as well as rote learning (and possibly more so), and the effect may be lasting as a total evaluation of qualitative learning.

Student attitudes indicated that M students perceived tests as assessments (negative) while Q students perceived tests as study aids (positive).

In summary, a theoretical implication of this study is that by increasing the frequency, precision, and cognitive level of value, the enhancement of learning is increased. Further, the specificity required in higher cognitive level multiple-choice questions may make them particularly useful instructional tools.

Interpretations

The authors of this research summarized their conclusions by saying that there is a growing body of emperical evidence that frequent questioning with immediate feedback significantly affects the quantity of material that is learned and retained, regardless of the nature of the subject matter. Less attention has been paid to the quality of learning that occurs as to the way in which systematic reinforcement affects that quality. Evidence shows that students who can answer essay questions are often unable to solve numerical-answer problems, whereas the reverse is rarely true. This research goes on to show that students can shift more easily from specificity to generalization than the opposite. If a multiple-choice question is well-written (free of ambiguity and triviality) it seems clear that higher cognitive level questions lead to greater learning and retention. Increasing frequency, precision, and cognitive level of testing and feedback to some as yet unspecified optimum level, causes the enhancement of learning by systematic reinforcement.

ANSTRACTOR'S ANALYSIS

As stated previously, the research of Professors Fisher, Williams and Roth contributes significantly to the present body of knowledge pertaining to how learning takes place and is retained. Strong evidence indicates that the learning process can be enhanced by precise, frequent and immediate feedback to the learner. All indications are that the CASE System prompts students to learn in small increments such

that the development of sound connections between new material and pre-existing knowledge is encouraged. Also, indications are that the CASE System has motivational value over traditional instructional methods. A case is also made for the benefactors of programmed/audio-tutorial/personalized instruction.

The study also indicated that attrition was unaffected by the difference of treatment groups. This is at least, partially due to the fact that both groups were exposed to identical instructional materials, the same teaching staff and similar one-to-one consultations. Although the bulk of this research did not investigate new and/or different strategies of instruction and learning enhancement, it should be mentioned that the study contributed to the validation of and verification of the significance of audio-tutorial education as a viable post-secondary tool.

One new conceptual contribution of this study is the evaluation of "meaningful" learning, defined by the authors as being that learning requiring higher cognitive skills and/or problem solving abilities. At this juncture, a majority of science educators feel that higher level cognition and its measurement cannot be evaluated by multiple-choice testing. The authors' research seems to indicate that, in fact, this prevailing opinion may not be true, and multiple-choice tests can be effective in promoting "meaningful" learning.

The validity of this study is well-founded, and the methodology and procedure as presented indicate no basis for doubt and/or question. Use of pre-test, post-test, varied treatment, and long-term retention testing are classic in educational research. The research design therefore, is sound and reproducible. Professors Fisher, Williams and Roth are to be commended for their work. It is especially rewarding to find that not only have they evaluated results and/or interpretations, but have also addressed and attempted to answer sources of criticism.

Fisher, Williams, and Roth should be encouraged to continue their research, and especially deal with larger numbers over greater periods of time. Future avenues of research building on their present data might include:

1) The use of more specifically delineated cognitive levels in their research.
2) Experiments with frequency of testing within the CASE System.
3) Retention evaluations over greater periods of time with special emphasis on comparisons and the establishment of norms.
4) Experimentation with more heterogenous grouping as might be found in general biology, general chemistry, and general physics. Such an effort would also allow for larger study numbers and validity.
5) Further attitudinal studies.

In summary, this research has contributed significantly to the body of knowledge of science education and/or learning.

REFERENCES

Ammons, R. B. "Effects of Knowledge of Performance: A Survey and Tentative Formulation." Journal of General Psychology, 54: 279-299, 1956.

Ausubel, D. P. Educational Psychology: A Cognitive View. New York: Holt, Rinehart, & Winston, 1968.

Bloom, B. S., Ed. Taxonomy of Educational Objectives, Handbook I: Cognitive Domain. New York: Davis McKay Co., Inc., 1956.

Gagne, R. M. The Conditions of Learning. New York: Holt, Rinehart, & Winston, 1970.

Keller, F. S. "Goodbye, Teacher . . ." Journal of Applied Behavioral Analysis, 1: 79-80, 1968.

Postlethwait, S. N. "A Systems Approach to Botany." Audiovisual Instruction, 8: 243-244, 1963.

Pressey, S. L. "Development and Appraisal of Devices Providing Immediate Automatic Scoring of Objective Tests and Concommitant Self-Instruction." The Journal of Psychology, 29: 417-447, 1954.

Skinner, B. F. "Teaching Machines." Science, 128: 969-977, 1958.

Skinner, B. F. "The Science of Learning and the Art of Teaching." Harvard Educational Review, 24: 86-97, 1954.

Larson, M. A., M. Forrest and L. Bostian. "Participation in Pro-Environmental Behavior." Journal of Environmental Education, 12 (3): 21-24, 1981.

Descriptors--*Affective Behavior; Behavior; *Citizen Participation; *Community Action; *Conservation Environment; *Environment; Environmental Education; Pollution; Surveys

Expanded abstract and analysis prepared especially for I.S.E. by Thomas R. Koballa, Jr., The University of Texas at Austin.

Purpose

The purpose of this investigation was to assess the universality of participation in pro-environmental behavior using large, random samples of the general United States population.

Rationale

The primary goal of environmental education is the development of citizenry capable of translating acquired knowledge about the biophysical environment and its related problems into pro-environmental behavior. Despite the difficulties associated with this less-than-deterministic transfer mechanism, annual reports indicate increased concern regarding the seriousness of air and water pollution.

While the reports indicate increased concern, the question of how much of this concern is transferred to pro-environmental behavior remains. Tognacci et al. (1972) claim that the increase in concern reflects only a growing commitment by a disproportionally small group of persons, causing little change in the amount of pro-environmental behavior. The extant data on pro-environmental behavior proves to be of little value in answering this question, having been gathered from select samples.

Research Design and Procedures

Data were gathered using two surveys; an exploratory survey of persons randomly selected from the Madison, Wisconsin, telephone directory and a statewide survey of Wisconsin residents 18 years of age and older. The surveys were administered during September 1972, and October 1974, respectively.

The exploratory survey questioned respondents regarding their participation in environmental activism in the political process and in home life using two scales; The Political Participation in Environmental Activism Scale and the Household Environmental Activism Scale. The scales were chosen because they contain items that describe activities over which most persons have control, such as letter writing, petition signing, non-use of paper plates, and use of returnable soft drink bottles. The two scales were mailed to a random sample of 400 persons living in Madison, Wisconsin. The final response rate was 57.5 percent or 230 households.

The statewide survey, conducted by personnel from the Wisconsin Survey Research Laboratory of the University of Wisconsin, Madison, sampled 544 adults in personal interviews. The survey consisted of five environmental activism questionnaire items that were operationalized in a manner so as to distinguish between general political and environmental activism. Representative of the questions that constituted the survey is the following item:

In the past year or so, have you signed a petition? (Yes) (No)

a. Did the petition relate to environmental issues? (Yes) (No)

b. Did the petition favor or oppose the environmentalist's point of view? (Favor) (Defend) (Oppose)

To provide data from a "known-group" to compare with the data collected from the statewide sample, the five environmental activism items were included on a questionnaire mailed to 207 environmental activists (i.e., those persons with Madison, Wisconsin, addresses on the mailing lists of various environmental organizations). Those sampled were assumed to have changed or directed their lifestyle to include more environmentally sound practices. The assumption that the "known-group" would as a whole respond more favorably than the general population toward pro-environmental behavior proved to be justified by the fact that 70 percent of the 164 respondents who returned the survey agreed with the statement, "I consider myself an environmental activitist."

Findings

The responses to the exploratory survey of persons randomly selected from the Madison, Wisconsin, telephone directory showed that approximately 30 percent of the respondents engaged in some form of political or household environmental activism. Subsequent analysis revealed that environmental activism was not restricted to a select sampling of respondents. Moreover, analysis revealed that persons who engaged in one or several of the political participation variables tended not to be active in household environmental activism. However, those individuals active in household activism variables were less likely to engage in environmental politics than those who were not involved in household environmental activism.

The data gathered as a result of the statewide survey suggested that at least one-third of the political activism in the state of Wisconsin during 1974 involved pro-environmental activism. Additional analysis of the results of the statewide survey showed that it was not a small group of persons who were participating in all of the pro-environmental behavior.

The data collected from the "known-group" of environmental activists showed that about 75 percent of this select sample reported participating in each of the five pro-environmental behaviors asked about.

Interpretations

The results of the exploratory and statewide surveys may be summarized as follows. As indicated by the exploratory survey, pro-environmental activism in the political process seems to be a better predictor of commitment to working to solve environmental problems than pro-environmental activism in the household.

From the results of the statewide survey, one can conclude that a large portion of the general population of Wisconsin was participating in political activism, with a little less than one-half of that devoted to pro-environmental activism. The high rate of pro-environmental response to the five item questionnaire when administered to the "known-group" of environmental activists suggest it to be a valid indicator of environmental activism. Subsequent use of the five-item questionnaire should prove useful in identifying environmentally concerned and active citizens.

ABSTRACTOR'S ANALYSIS

It has been suggested that the increase in environmental concern reflects only a growing commitment by a minute segment of the total population (Tognacci et al., 1972). This suggestion has neither been proven nor disproven, mainly because of extant data having been collected from select samples. Random sampling of the general population is offered by the investigators as a way of grappling with this basic question.

This paper presents a rather scanty review of the literature regarding environmental concerns and pro-environmental behavior. From the outset, the rationale for the study is undergirded by the findings of only two prior studies, one being marginally related to the question under investigation.

The exploratory survey of persons randomly selected from the Madison telephone directory suffers from several flaws. To begin with, the population of Madison, Wisconsin, can hardly be characterized as typical of the general population of the United States. Madison, the capital of Wisconsin, is a university town with little hard industry. It is inhabited by relatively affluent, highly educated, moderate to liberal persons. Being the capital of Wisconsin, Madison residents are afforded opportunities beyond those typically found elsewhere to become involved in environmental politics. Furthermore, Madison provides its residents with opportunities to recycle cans and newspapers with minimal effort.

With a final response rate ot 57.5 percent, the results of the survey can hardly be said to adequately reflect the views of the general population, let alone the citizens of Madison, Wisconsin. It is quite possible that the conclusions put forth in the study were based on data collected from a less than representative sample. Persons venting their animosity toward environmentalists or their efforts could very well represent a large proportion of the 42.5 percent who did not return the mailed survey.

Furthermore, the finding that persons who engaged in one of the political participation variables were not active in their households seems to contradict the self-persuasion literature reported by social psychologists and speech communicators. The self-persuasion literature suggests that by involving oneself in activities intended to persuade others of the need for pro-environmental behavior should be sufficient to induce dissonance in the mind of the persuadee not involved in the behavior advocated. Dissonance may be reduced by persuadee involvement in environmental activism, possible at home.

The results of the statewide survey of Wisconsin residents 18 years of age and older did not suffer from the many flaws that plagued the exploratory survey. Interviewing persons identified by stratified random sampling aided in reducing the likelihood of response bias normally associated with mailed surveys. Also, the five-item questionnaire employed in data collection can be viewed as a valid identifier of environmentally concerned and active citizens based on the results of its use with a known-group of environmental activists. Household environmental activism was not addressed by this five-item questionnaire.

A most distressing factor of this work is the approximate 10 year lapse between data collection and reporting. Behaviors associated with environmental activism have changed since 1972. For example, items on the Household Environmental Activism Scale inquire about respondents' use of non-lead gasoline and non-phosphate laundry detergent are no longer applicable. More recently these behaviors have become standard household practice resulting from government legislation.

If further information is sought regarding the status of environmental activism, better sampling techniques must be employed. The use of better sampling techniques should permit the generalizability of reported findings to the total United States population.

REFERENCE

Tognacci, L. N., R. H. Weigel, M. F. Wideen and D. T. Vernon. "Environmental Quality: How Universal is Public Concern?" Environment and Behavior, 4: 72-86, 1972.