DOCUMENT RESUME

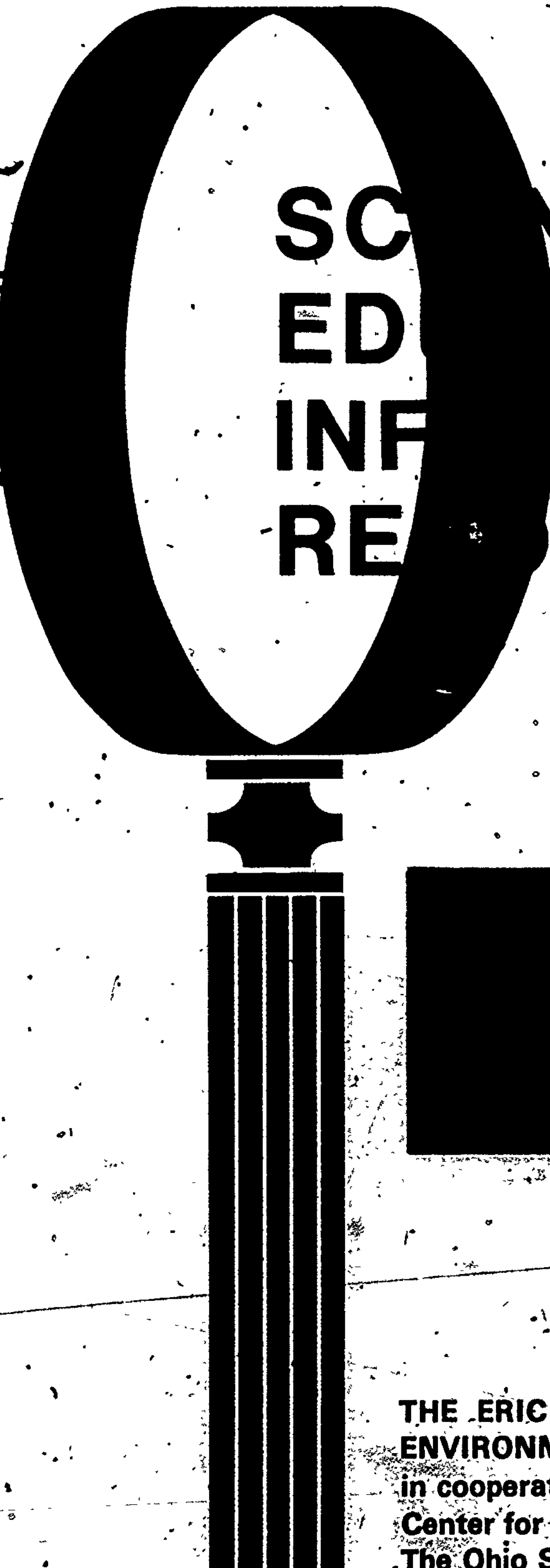
ED 216 891 SE 037 813

AUTHOR Blosser, Patricia, E., Ed.; Mayer, Victor J., Ed.
TITLE Investigations in Science Education. Volume 8, Number 2.
INSTITUTION ERIC Clearinghouse for Science, Mathematics, and Environmental Education, Columbus, Ohio.; Ohio State Univ., Columbus. Center for Science and Mathematics Education.
PUB DATE 82
NOTE 84p.
AVAILABLE FROM Information Reference Center (ERIC/IRC), The Ohio State Univ., 1200 Chambers Rd., 3rd Floor, Columbus, OH 43212 ($1.75 single copy; $6.00 per year subscription price).
JOURNAL CIT Investigations in Science Education; v8 n2 1982

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Cognitive Development; *College Science; Computer Assisted Instruction; Concept Formation; Educational Research; Elementary Secondary Education; Higher Education; Reading Comprehension; Reading Readiness; Science Education; *Science Instruction; Science Laboratories
IDENTIFIERS *Science Education Research

ABSTRACT

Volume 8, No. 2 of Investigations in Science Education contains critiques of articles dealing with science education research about various aspects of instruction. Also included are five responses from authors of articles to critiques of their work as it has been reported in professional journals. In total, this volume contains 10 critiques and five responses to critiques. The student differences, instruction and reading comprehension, relationship of curriculum materials to student self-concept, the use of concrete exemplars in teaching chemistry, the effectiveness of computer simulated laboratory experiences, student cognitive structure and the structure of science concept, and the ability to distinguish between teleological and causal explanations. (PEB)



SCIENCE EDUCATION INFORMATION REPORT

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

ERIC® Clearinghouse for Science, Mathematics, and Environmental Education

INVESTIGATIONS IN SCIENCE EDUCATION

Volume 8, Number 2, 1982

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: $6.00 per year. Single Copy Price: $1.75
Add 50¢ for Canadian mailings and $1.00 for foreign mailings.

INVESTIGATIONS IN SCIENCE EDUCATION

Volume 8, Number 2, 1982

5

NOTES FROM THE EDITOR

This issue has as its major theme the analysis of articles dealing with various aspects of instruction. Sheehan and Hambleton looked at adapting instruction to student differences. Thomas looked at methods for increasing students' reading comprehension of textbooks. Anderson and Fowler investigated the effects of using behavioral objectives. Hermann and Hincksman compared two approaches to teaching chemistry. Krockover and Malcolm investigated the use of specific curriculum materials on the child's self-concept.

The analyses found in the "Critiques and Responses" section also deal with instruction. Goodstein and Howe studied the use of concrete exemplars in teaching chemistry. Quorn and Yore compared the effects of dissimilar programs on reading readiness of kindergartners. Cavin and Lagowski studied the effectiveness of computer simulated laboratory experiences in a college chemistry course. Preece compared the use of two models of the structure of physical science concepts on graduate students' cognitive structure of these concepts. Bartov investigated whether or not students could be taught to distinguish between teleological and causal explanations.

These articles present evidence that there are many variables to be considered when planning for teaching and when choosing instructional methods to use in the science classroom.

Patricia E. Blosser
Editor

Victor J. Mayer
Associate Editor

INSTRUCTION

Sheehan, Daniel S. and Ronald K. Hambleton. "Adapting Instruction to Student Differences in an Individualized Science Program." Journal of Research in Science Teaching, 14(1): 27-32, 1977.

Descriptors--*Achievement; *Educational Research; Grade 9; *Individualized Instruction; *Instruction; *Predictor Variables; Science Education; *Science Courses; *Secondary School Science

Expanded abstract and analysis prepared especially for I.S.E. by G. Neufeld, Brandon University.

Purpose

This study was designed to investigate aptitude-treatment interactions (ATI)--whether grade nine students with certain aptitudes learn more science when taught using one teaching method than similar students learn from other teaching methods.

Rationale

Many recent educational models, such as Individually Prescribed Instruction (Glaser, 1968), Project Plan (Flanagan, 1967), and a Model of School Learning (Carroll, 1963), emphasize adapting instruction to the students' interests, abilities, aptitudes, and needs. Over the past 15 years there have been many studies of aptitude-treatment interactions (Cronbach and Snow, 1969, 1977; Snow, 1976; Tobias, 1976; Witkin et al., 1977). In general, there have been few solidly demonstrated ATI effects. However, the authors feel that the following tentative principles for the design of instruction can be inferred from the literature:

1. Instructional treatments which reduce the burden of semantic processing of verbal information (such as programmed instruction or a media-oriented approach) should decrease the relationship between general ability and achievement.
2. Instructional treatments which place more responsibility on the student for organizing material should strengthen the relationship

between achievement and such aptitudes as general ability, achievement, motivation, study habits, attitudes, and anxiety.[3]

The authors used these principles as the basis for designing the instructional treatments and for selecting the student aptitudes to be measured. As a result, the study serves as an empirical test of the validity of these principles.

Research Design and Procedures

The research design used in this study is shown in Figure 1. The design is a variation of the Pretest-Posttest Control Group design

$O_1 -- R -- T_1 -- O_2$	O_1 = battery of aptitude measures and pretest
$O_1 -- R -- T_2 -- O_2$	O_2 = criterion-referenced posttest
$O_1 -- R -- T_3 -- O_2$	R = random assignment to treatments
$O_1 -- R -- T_4 -- O_2$	T_{1-4} = four different instructional treatments

Figure 1: Schematic representation of the research design

(Campbell and Stanley, 1963) in that there are multiple treatment groups but no control group.

The subjects for the study were 285 students enrolled in a ninth-grade science program. The report does not indicate the general ability level of the students, their science backgrounds, their socioeconomic status, or whether they even attended the same school. The students were randomly assigned to one of the four treatment groups. There is no indication that the random assignment was stratified.

Each student wrote a total of 12 aptitude tests prior to random assignment to the various instructional treatment groups. The aptitude tests used were:

1. Module Pretest (40 four-alternative multiple-choice questions --no reliability or validity data).

*2. Lorge-Thorndike Intelligence Test.

*3. Science Research Associates (SRA) Achievement Series.

*4. Survey of Study Habits and Attitudes.

*5. Junior Index of Motivation (JIM Scale).

*6. Test Anxiety Scale for Children.

*7. Children's Manifest Anxiety Scale.

*8. Intellectual Achievement Responsibility Scale.

*9. School Anxiety Scale.

10. Mathematics Test (no reliability or validity data) (French, Ekstrom and Price, 1963).

11. Letter Sets Test (no reliability or validity data) (French et al., 1963).

12. Student Attitude Questionnaire (semantic differential format --no reliability or validity data).

Most of these aptitude test instruments (those marked with an *) are standardized or quasi-standardized and reliability and/or validity data are available in the literature. The module pretest and the student attitude questionnaire were devised by the researchers and no validity or reliability data are provided.

The instructional treatments were described in the study as:

Teacher-Directed Treatment. Students assigned to this instructional treatment met in a teacher-led class and were presented a series of lectures by ninth-grade science teachers.

Media Treatment. Students in this treatment worked alone or in pairs on four worksheets that were used in conjunction with one video tape and seven audio tapes.

Reading Treatment. This treatment consisted of a reading handout composed of an assignment section and a section of appropriate readings. Each student who was assigned to this instructional treatment worked alone on his booklet of readings.

Programmed Instruction Treatment. Students assigned to this instructional treatment worked alone on one set of five booklets consisting of 185 frames of programmed instruction material.

There is no indication of what time of year the study was conducted, its duration, or whether the students in the various instructional treatment groups received the same number of hours of instruction.

The criterion measure for the study was a Module Posttest. It consisted of 40 four-alternative multiple-choice questions. No validity or reliability data are provided. The report does not indicate if this posttest was identical to the pretest given prior to instruction (both have 40 items based on the module objectives). The Module Posttest was administered immediately after instruction and again a month later (Delayed Module Posttest).

The data were analyzed using a regression analysis. The criterion variable (score on immediate or delayed posttest) was regressed on the various aptitude variables for each treatment group. Unspecified statistical procedures were then used to test the parallelism of slopes of the different groups for each aptitude measure. Due to technical problems with the media treatment, the analyses were repeated excluding this group.

In cases where the parallelism of slopes hypothesis was rejected, a Johnson-Neyman analysis (Potthoff, 1964) was performed on each pair of treatment groups within each interaction to determine the utility of the interaction.

Findings

In three cases the parallelism of slopes hypothesis was rejected at the 0.05 level. The Johnson-Neyman analysis indicated that in only one of these cases was the interaction disordinal. This occurred when the Test Anxiety Scale for Children was the aptitude, the Delayed Module Posttest was the criterion, and the media treatment group was excluded. Only the reading and programmed instruction treatments had significantly different slopes at the 0.05 level. The Delayed Module Posttest scores rose slightly with increased test anxiety scores for the reading treatment (slope ≈ + 0.04) and fell fairly rapidly for the programmed instruction treatment (slope ≈ - 0.68). The Johnson-Neyman region of nonsignificance for these two groups lay between scores of 5.8 and 23.3 on the Test Anxiety Scale for Children. This indicates that students with scores below 6 should be assigned to the programmed instruction treatment and those with scores above 23 should receive the reading treatment.

Interpretations

The authors feel that their results should be interpreted with considerable caution because of the possibility that the three interactions that were detected were simply chance results. They feel that the lack of definitive results was due to faulty design of the study--specifically with the construction and execution of the instructional treatments. It appeared that the treatments were not separated as completely as would have been desirable for experimental purposes.

ABSTRACTOR'S ANALYSIS

The study of aptitude treatment interactions (ATI's) is a relatively new field of research. It began about 20 years ago in response to Cronbach's (1957) call for cooperative research between experimental and correlational psychologists.

This new area of investigation has generated a great deal of interest because the differential instruction of individuals based on their needs, interests, aptitudes and abilities is the cornerstone of individualized instruction. The numerous studies of ATI's have been capably reviewed on several occasions (Cronbach and Snow, 1969; Berliner and Cohen, 1973; Tobias, 1976; Cronbach and Snow, 1977). Despite the many studies in this area, progress has been slow. In general, the research reviews indicate that there are few, if any, replicated ATI's that permit prescriptive instruction.

In view of this discouraging conclusion reached by the various reviewers of the research on ATI's, it is not clear how this study fits into the matrix of previous research in this field or on what basis the authors inferred the tentative guidelines for the design of instruction that form the basis for the study.

In their reviews of the research in ATI's, all of the reviewers have commented on a common weakness of ATI studies that affects their validity: that the aptitude variables and the corresponding instructional treatments have not been very carefully thought out. Often the instructional treatments have differed only in minor details so that these treatments have not tapped different aptitudes. In view of this continuing criticism it is unfortunate that Sheehan and Hambleton provide such a brief and cursory description of the instructional treatments used in the study. A more extended description would have allowed a reader to assess the validity of the authors' statement that the paucity of results was due in large part to problems in the construction of the instructional treatments.

The validity of the study is further weakened by the relatively short duration of the instructional treatments. The authors do not mention exactly how long the treatments lasted but treatment brevity can be inferred from their statement that the treatments extended over a small segment (module) of a ninth-grade science course. Practicing teachers are well aware of the fact that an abrupt change in teaching methodology, such as from a lecture-lab method to programmed instruction or a set of readings and worksheets, can serve as a strong motivator but can also generate a great deal of confusion over a several week period as the students and teachers gradually adapt to their new roles and responsibilities. As a result, any potential differences in the effectiveness of the various treatments on different students may well have been swamped by these extraneous factors.

The research design chosen for the study was an unusually rigorous one. Undoubtedly the researchers encountered stiff opposition to the random assignment of students to treatments and they are to be commended for insisting on its use--it certainly strengthens the validity of the study.

In general, the research report is well written. The candor with which they discuss possible reasons for their failure to detect more ATI's is refreshing. However, they fail to mention in their discussion any consideration of the power of the statistical tests they employed. (Statistical power is the probability of detecting significant relationship when the relationship does, in fact, exist.) A number of statisticians, Cohen (1962), Brewer (1972), and Schmillain (1976), have commented on the very low statistical power inherent in most educational research. Educational researchers planning studies would be well advised to keep abreast of the literature on hypothesis testing and power analysis. Cohen's book, Statistical Power Analysis for the Behavioural Sciences (Rev. ed., 1977), clearly outlines the concepts and procedures involved and should be on every researcher's bookshelf. Cohen's (1979) proposal for new conventions for designing and reporting research results deserves widespread attention from the research community.

The research report would have been more useful if it had more complete description of how the tentative guidelines for the design of the treatments were related to previous studies. A description of the previous background and socioeconomic status of the subjects would have allowed the reader to gauge the generalizability of the results. A description of how the battery of aptitude tests was administered would have allowed the reader to assess the possible effects of fatigue and test weariness on the aptitude test scores. As previously mentioned, a more adequate description of the instructional treatments and their duration would have allowed the reader to assess whether the treatments engaged different aptitudes and whether novelty effects could have swamped the ATI.

The field of ATI's has been, and still is, an important and popular one. However, as the previously mentioned research reviews have pointed out, the problems are many and progress has been agonizingly slow. Future researchers in this area must be careful to design their instructional treatments so that they do engage different aptitudes, to ensure that the treatments extend over sufficient time so that the ATI's are not swamped by extraneous factors, and, to design their studies with sufficient statistical power so that the hypothesized effects are likely to be detected.

REFERENCES

Berliner, D. C. and L. S. Cohen. "Trait-Treatment Interaction and Learning." Review of Research in Education 6: 58-94, 1973.

Brewer, J. K. "On the Power of Statistical Tests in the American Educational Research Journal." American Educational Research Journal 9: 391-401, 1972.

Campbell, D. T. and J. C. Stanley. Experimental and Quasi-Experimental Designs for Research (reprinted from the Handbook of Research on Teaching). Chicago: Rand McNally & Co., 1963.

Carroll, J. B. "A Model of School Learning." Teachers College Record 64: 723-733, 1963.

Cohen, J. "The Statistical Power of Abnormal-Social Psychological Research: A Review." Journal of Abnormal and Social Psychology 65: 145-153, 1962.

Cohen, J. Statistical Power Analysis for the Behavioral Sciences (Rev. ed.). New York: Academic Press, 1977.

Cohen, S. A. and J. S. Hyman. "How Come So Many Hypotheses in Educational Research are Supported? (A Modest Proposal)." Educational Researcher 8(11): 12-16, 1979.

Cronbach, L. J. "The Two Disciplines of Scientific Psychology." American Psychologist 12: 671-684, 1957.

Cronbach, L. J. and R. E. Snow. Individual Differences in Learning Ability as a Function of Instructional Variables (Final Report, Contract No. OEC-4-6-061269-1217, U.S. Office of Education). Stanford, Calif.: Stanford University, School of Education, 1969. ED 029 001.

Cronbach, L. J. and R. E. Snow. Aptitudes and Instructional Methods: A Handbook for Research on Interactions. New York: Irvington/Naiburg, 1977.

Flanagan, J. C. "Functional Education for the Seventies." Phi Delta Kappan 49: 27-32, 1967.

French, J., R. Ekstrom, and L. Price. Manual for Kit of Reference Tests for Cognitive Factors. Princeton, N.J.: Educational Testing Service, 1963.

Glaser, R. "Adapting the Elementary School Curriculum to Individual Performance." In Proceedings of the 1967 Invitational Conference on Testing Problems. Princeton, N.J.: Educational Testing Service, 1968.

Potthoff, R. F. "On the Johnson-Neyman Technique and Some Extensions Thereof." Pyschometrika 29: 241-256, 1964.

Schmelkin, L. "Statistical Power Analysis of Research in Exceptional Children." Paper presented at the 54th Annual International Convention of the Council for Exceptional Children, Chicago, April 1976. ED 126 639.

Snow, R. E. "Research on Aptitude for Learning: A Progress Report." In L. S. Shulman (Ed.), Review of Research in Education (Vol. 4), Itasca, Ill.: F. E. Peacock, 1976.

Tobias, S. "Achievement Treatment Interactions." Review of Educational Research 46: 61-74, 1976.

Witkin, H. A., C. A. Moore, D. R. Goodenough, and P. W. Cox. "Field-Dependent and Field-Independent Cognitive Styles and Their Educational Implications." Review of Educational Research 47: 1-64, 1977.

Thomas, J. L. "The Influence of Pictorial Illustrations with Written Text and Previous Achievement on the Reading Comprehension of Fourth Grade Science Students." Journal of Research in Science Teaching 15: 401-405, 1978.

Descriptors--*Academic Achievement; Elementary School Science; Elementary School Students; *Grade 4; *Illustrations; *Reading Comprehension; *Science Education; Textbooks

Expanded abstract and analysis prepared especially for I.S.E. by William G. Holliday and William Winn, The University of Calgary.

Purpose

Dr. Thomas asked three research questions. First, do the addition of line drawings or photographs adjunct to a textual description of unfamiliar science information facilitate, in fourth grade students, comprehension of the verbally expressed information? Second, do the composite SRA reading and science achievement scores of these learners predict "literal" and "inferential" comprehension of unfamiliar science information? Third, what is the predictive relationship between reading comprehension and general science achievement, as measured by SRA assessment instruments?

Rationale

Publishers of elementary school science materials use many different types of line drawings and photographs to illustrate ideas described in the prose portion of their textbooks. Yet, few investigators have evaluated the learning effectiveness of these pictures in terms of previous achievement, and colored-uncolored and inclusion-exclusion pictorial variables. Previous research suggests that pictorial illustrations either have no effect on comprehension (Samuels, 1970) or have unclear effects on comprehension (Holliday, 1973, 1975). Indeed, Travers and Alvarado (1970) indicate that publishers make editorial decisions about including textbook pictures based on unsystematic intuitive reasons rather than on empirical grounds, hence, the need to investigate in a general fashion the learning effects of pictures adjunct to prose passages.

Research Design and Procedure

A sample of 108 fourth-grade students from three elementary schools was randomly assigned to a multiple colored photograph-text, single colored line drawing-text or a text-only treatment. The first treatment consisted of pictorial and textual displays taken directly from a recently published book. The second treatment was a simplified (apparently uncolored) drawing substituted for the publisher's photograph and set in a single colored background. The third treatment consisted of the publisher's prose description without illustrative adjunct pictures. These experimental materials were photographed and projected on a screen using a two-by-two slide projector. Subsequently, students were administered a comprehension posttest. This procedure was conducted on two occasions, (apparently) attending to two science topics at each sitting. In addition, subjects were administered the SRA reading and science achievement tests for use in evaluating the second and third research questions.

The four science topics chosen from a group of 20 were identified by the subjects as those least familiar to them. In addition, the experimenter used Fry's readability graph to assess the appropriateness of the chosen prose material and used Smith and Barrett's (1974) taxonomy of reading comprehension in the development of the posttest to insure that items measured either literal (i.e., recognition) or inferential (e.g., implicit main ideas) learning.

Findings

No significant differences were found among the group mean scores nor among treatment-by-SRA test data. However, significant differences in posttest scores were detected in the predicted direction among the high, medium and low performers (i.e., high > medium > low performers) on the composite score variable (SRA reading comprehension and science achievement). Finally, analysis of the SRA composite score yielded a positive correlational coefficient (0.85), as predicted.

Interpretations

It was conditionally concluded that the type of adjunct picture (drawing vs. photograph) and its inclusion or exclusion apparently had no influence on student comprehension of science information. In this regard, "Publishers could greatly reduce some of the enormous expense of reproducing and including color photographs" in their science textbooks. Furthermore, these books apparently were written for the average student when, indeed, such averaging failed to meet the learning needs of students of high or low abilities.

ABSTRACTOR'S ANALYSIS

This study did not contribute to our understanding about how visual media facilitates in children comprehension of science information presented adjunct to a prose passage. Furthermore, the correlational findings dealing with science comprehension and the SRA instruments were not scientifically significant. Indeed, Dr. Thomas was cognizant of weaknesses contained in his study, as suggested by his remarks to us and by those made to his audiences (Thomas, Note 1) at the 1977 Association for Educational and Communications and Technology (AECT) annual meeting. Based on his comments and other apparent abilities, we believe he is capable of making a substantial contribution to the field of educational communication. Nevertheless, we will describe some of the more serious problems associated with the present study and cite recent methodological advances and empirical findings outlined by other researchers exploring the learning effects of adjunct visuals.

The most serious problem constitutes Dr. Thomas' pointed suggestion that "pictures do not influence comprehension." Careful inspection of other research studies does not substantiate this "no-effect" suggestion and the experimental results used to explore this general hypothesis do not provide the evidence needed to give it reasonable support for three reasons. First, a single experiment providing "no differences" or negative evidence does not constitute grounds for suggesting that two or more instructional treatment conditions have a similar effect on learners. In fact, such findings are of little scientific interest and are often considered meaningless when an experimenter theoretically deduces from the literature an expectation of positive results. In this sense, the literature review presented in the article is incomplete and somewhat misleading. For example, Holliday's (1975) and Holliday and Harvey's (1976) work clearly supports the general hypothesis that adjunct pictures can facilitate comprehension in secondary school science students. Yet, reading Dr. Thomas' literature review, one is left with the impression that this hypothesis lacks empirical support.

Second, Dr. Thomas chose not to use a traditional control group, thereby leaving open to question the possibility that subjects did not comprehend on the average treatment information. Data generated from a control group would have provided a baseline of performance from which experimental data could have been interpreted. Indeed, interpretation of any kind is most difficult (if not impossible) under such "no-control" conditions when no significant differences are reported among treatment groups. The checklist mechanism used in this study to select treatment information cannot be considered an adequate control of previous knowledge because of the unreliability of such assessment methods.

Third, the experimental variables are vaguely defined. Specifically, the following variables are not described in the article: 1) the "science" topics presented to the children, 2) an operational definition of the prose passages and the "comprehension" posttest, 3) the informational nature of the pictures and the degree of prose-picture information overlap, and 4) a clear description of the procedures used to treat and assess the subjects. (Incidentally, some of this information was presented in Dr. Thomas' AECT address.) On the other hand, Dr. Thomas' well-written article suggests (but does not confirm) that a conscientious effort was made to provide fourth-grade children with "appropriate" learning materials and "reasonable" posttest items.

Without question, pictures can facilitate the learning of prose material, contrary to earlier claims made by such researchers as Vernon (1953) and Samuels (1970). Most of the recent work in this area (see Holliday, 1973; Levin and Lesgold, 1978) has evaluated elementary school children listening to fictional narratives read to them while viewing a line drawing illustrating the main ideas of the story. Those studies taken together support this "pictures-can-help" hypothesis and permit some generalization across: 1) methods of presentation (e.g., verbal information presented in isolated sentences and in prose forms and visual information displayed in booklets projected on large classroom screens, and presented to children as laminated plastic cutouts), 2) methods of testing (i.e., cued recall using verbatim and paraphrased questions), 3) learner aptitudes (i.e., sex, age, social class, intellectual ability), 4) passage characteristics (i.e., length, complexity, topic of narrative), and 5) retention time (i.e., immediate and delayed recall).

Pictorial research in science education during the past decade has mainly focused attention on junior and senior high school students learning "textbook" information using adjunct line drawings and photographs (Dwyer, 1972, 1978; Holliday, 1975; Holliday and Harvey, 1975) and specialized learning materials including flow diagrams (Holliday, 1976; Winn, in press a). On the whole, the empirical evidence and theoretical explanations suggest that visual media can facilitate science learning. However, whether pictures have a similar effect on children learning science processes and concepts is currently a matter of speculation. Consequently, science educators interested in this research area will be pleased to learn that recently reported methodological procedures (see Levin, Bender and Pressley, 1979) can be used with confidence to examine a multitude of adjunct picture hypotheses dealing with elementary school science. In addition, theoretical frameworks used to deduce research questions about the learning effects of pictures on young science students are derivable from commonly cited Piagetian sources and a wide variety of recent non-Piagetian works (e.g., Ausubel, Novak and Hanesian, 1978; Brown and Smiley, 1978; Gagné and White, 1978; Pressley, 1977; Schallert, 1978; Winn, in press b).

REFERENCE NOTE

1. Thomas, J. L. "The Influence of Pictorial Illustrations with Written Text and Previous Achievement on the Reading Comprehension of Fourth Grade Science Students." Paper presented at the annual meeting of the Association for Educational Communication and Technology, Kansas City, April 1978.

REFERENCES

Ausubel, D. P.; J. [illegible] and H. Hanesian. Educational Psychology: A Cognitive Vi[illegible]: Holt, Rinehart and Winston, 1978.

Brown, A. L. and S. S. Smiley. "The Development of Strategies for Studying Texts." Child Development 49: 1076-1088, 1978.

Dwyer, F. M. "The Effects of Overt Responses in Improving Visually Programmed Science Instruction." Journal of Research in Science Teaching 9: 47-55, 1972.

Dwyer, F. M. Strategies for Improving Visual Learning. State College, Pennsylvania: Learning Services, 1978.

Gagné, R. M. and R. T. White. "Memory Structures and Learning Outcomes." Review of Educational Research 48: 187-222, 1978.

Holliday, W. G. "Critical Analysis of Pictorial Research Related to Science Education." Science Education 5: 201-214, 1973.

Holliday, W. G. "The Effects of Verbal and Adjunct Pictorial Information in Science Instruction." Journal of Research in Science Teaching 12: 77-83, 1975.

Holliday, W. G. and D. A. Harvey. "Adjunct Labelled Drawings in Teaching Physics to Junior High School Students." Journal of Research in Science Teaching 13: 37-43, 1976.

Holliday, W. G. "Teaching Verbal Chains Using Flow Diagrams and Texts." AV Communication Review 24: 63-77, 1976.

Levin, J. R. and A. M. Lesgold. "On Pictures in Prose." Educational Communication and Technology Journal 26: 233-243, 1978.

Levin, J. R.; B. G. Bender; and M. Pressley. "Pictures, Imagery and Children's Recall of Central Versus Peripheral Sentence Information." Educational Communication and Technology Journal 27: 89-95, 1979.

Pressley, M. "Imagery and Children's Learning: Putting the Picture in Developmental Perspective." Review of Educational Research 47: 585-622, 1977.

Samuels, S. J. "Effects of Pictures on Learning to Read, Comprehension and Attitudes." Review of Educational Research 40: 397-407, 1970.

Schallert, D. L. "The Role of Illustrations in Reading Comprehension." In R. J. Spiro, B. C. Bruce, and W. F. Brewer (Eds.), Theoretical Issues in Reading Comprehension. Hillsdale, N.J.: Erlbaum, 1978.

Smith, R. J. and T. C. Barrett. Teaching Reading in the Middle Grades. Reading, Mass.: Addison-Wesley Publishing, 1974.

Travers, R.M.W. and V. Alvarado. "The Design of Pictures for Teaching Children in Elementary School." AV Communication Review 18: 47-63, 1970.

Vernon, M. D. "The Value of Pictorial Illustrations." British Journal of Educational Psychology 23: 171-179, 1953.

Winn, W. "The Effect of Block-Word Diagrams on the Structuring of Science Concepts as a Function of General Ability." Journal of Research in Science Teaching, in press. a.

Winn, W. "Visual Information Processing: A Pragmatic Approach to the Imagery Question." Educational Communication and Technology Journal, in press. b.

Anderson, Elaine J. and H. Seymour Fowler. "The Effects of Selected Entering Behaviors and Different Cognitive Levels of Behavioral Objectives on Learning and Retention Performance in a Unit on Population Genetics." Journal of Research in Science Teaching 15(5): 373-379, 1978.

Descriptors--*Behavioral Objectives; Behavior Patterns; Biology; *Critical Thinking; *Genetics; *Learning; *Preservice Education; Science Education

Expanded abstract and analysis prepared especially for I.S.E. by David R. Stronck, University of Victoria.

Purpose

The purpose of the study was to expand research information regarding the effects of selected entering behaviors and different cognitive levels of behavioral objectives on learning and retention performance in a unit on population genetics. The selected entering behaviors were prior knowledge and critical thinking. The different cognitive levels of behavioral objectives were (1) low, (2) high, (3) both low and high, and (4) none. The learning and retention performance was measured by criterion tests completed by 121 preservice elementary education majors. The unit of population genetics was a 97-frame excerpt from Population Genetics: A Self-Instructional Program by R. Anderson, V. E. Drantz, G. W. Faust, and J. T. Guthrie.

Rationale

Bloom (1956) edited the Taxonomy of Educational Objectives, Handbook I: Cognitive Domain. Many studies have used the clear definitions for classifying questions or behavioral objectives which this Taxonomy has provided. Some recent studies have demonstrated that learners function differently when they are challenged by different levels of the cognitive domain (Madaus, Woods and Nuttall, 1973; Anderson, DeMelo, Szabo and Toth, 1974). Science educators generally recommend an emphasis on the higher levels of the cognitive domain. Another major recommendation for the improvement of instruction arises from Ausubel's study (1963) of advance organizers. This study used behavioral objectives of different levels of the cognitive domain as advance organizers for a self-instructional program. Previous research studies encourage the anticipation that advance organizers will be most effective when they are in the higher levels of the cognitive domain.

Research Design and Procedure

This study used the pretest-posttest control group design. The pretest determined prior knowledge of basic genetic concepts (PKT). This pretest was given one week prior to the treatment. At this same time the 121 subjects completed the Cornell Critical Thinking Test (CCTT). Two posttests were given: (1) postcriterion test (PT_1) immediately after the treatment, and (2) postcriterion retention test (PT_2) eight weeks after the treatment.

Within the treatment was the study of the 97-frame excerpt from Population Genetics: A Self-Instructional Program. All subjects used the same 97-frame excerpt and proceeded independently at their own pace for three days. The first page of the learning packets was of four different types and therefore generated four different treatments: Treatment 1 was a list of low cognitive level behavioral objectives; Treatment 2 was a list of high cognitive level behavioral objectives; Treatment 3 was a list of both low and high cognitive level behavioral objectives; Treatment 4 was a placebo rather than objectives. Although this last list was made to appear as the lists of objectives, it consisted of a list of statements of genetic curiosities. Because the last treatment was not a treatment with behavioral objectives, the subjects receiving this treatment served as the control group.

The 121 subjects were preservice elementary education majors at a major university in Pennsylvania. They were divided into two equal groups on the basis of the prior knowledge test (PKT) by using a median split at the score of 19. Equal numbers of subjects with high or low prior knowledge of selected genetic concepts were randomly assigned to each of the four treatments. Similarly the Cornell Critical Thinking Test (CCTT) was used to divide the same subjects into two groups by using a median split at the score of 30. Equal numbers with high or low critical thinking scores were randomly assigned to each treatment. Combinations of subjects by scores on PKT and CCTT generated these four groups: (1) low scores on both tests, (2) low on PKT but high on CCTT, (3) high on PKT but low CCTT, and (4) high scores on both tests. Because each of these four groups were divided by four different treatments, 16 distinct subgroups were finally identified. The number of subjects in each of these final subgroups ranged from 3 to 11.

The data on the two postcriterion tests (one showing immediate learning performance and the other demonstrating retention performance after eight weeks) were analyzed by using a two-way analysis of variance to compare the four treatments and two levels of entering behaviors. Because the number of subjects in the subgroups were unequal, the program ANOVUM was selected from The Pennsylvania State University Computer Center library. Comparisons of the specific subgroups were made as well as the considerations of larger groupings within the study.

The postcriterion test consisted of 32 items considering the learning task on population genetics. The postcriterion test (PT_1) was given eight weeks later as (PT_2) with randomly reordered test items. These postcriterion tests had three subscales, each designed to measure a specific level of cognitive ability congruent with the behavioral objectives and the learning packets. The three subscales were for measuring knowledge, comprehension, and application which are the first three levels of Bloom's taxonomy of the cognitive domain. A selected panel of experts had 95 percent agreement level in establishing the content validity of each assessment tool used in the study.

Findings

The subjects in the four treatments did not differ significantly on posttest scores for either PT_1 or PT_2. On PT_1 differences were found between subjects with high and subjects with low prior knowledge levels when they experienced Treatment 3 (which consisted of both low and high cognitive levels of behavioral objectives). In this situation the group with high prior knowledge had significantly higher scores on the knowledge level questions. On PT_2 differences were found between subjects with high and subjects with low prior knowledge levels when they experienced Treatment 2 (which consisted of high cognitive levels of behavioral objectives). In this situation the group with high prior knowledge had significantly higher scores on both the knowledge and the comprehension questions. No significant differences were found on the basis of identifying the subjects as high or low critical thinkers.

Interpretations

The investigator made the following conclusions: "(1) Behavioral objectives meaningfully presented to some learners with a new cognitive task can enhance learning at the knowledge and comprehension levels. (2) Behavioral objectives based on cognitive levels of learning according to Bloom (1956) and consistent with programmed materials providing low and high cognitive level experiences enhanced learning for students identified as having high prior knowledge. (3) The identification of specific entering behaviors seems to be a significant variable for a learner's performance in content-specific learning tasks. (4) Systematically written programmed instruction seems to enhance learning for some students. (5) No interaction was noted between critical thinking ability and treatments. This leads one to the notion that the domains of critical thinking ability and the levels of cognitive behavioral objectives presented in this study are in fact independent of each other." The investigator observed that this study encourages college science teachers to preassess students' knowledge.

ABSTRACTOR'S ANALYSIS

Previous research studies suggested this study by which behavioral objectives of different levels of the cognitive domain were used as advance organizers. Researchers generally will anticipate that advance organizers of the higher levels will serve best as factors which improve learning. Research on testing indicated that students will be best prepared for any examination by anticipating essay questions which will require higher levels of the cognitive domain. Ausubel and others have demonstrated that advance organizers do improve learning. This study attempted to provide evidence in support of generally accepted concepts.

The design and procedure of this study seem correct, with the following exceptions: (1) All subjects proceeded independently at their own pace for three days to complete the learning from the 97 frames. (2) Although the total number of subjects was 121, the subgroups consisted of an average of only 7.56 students with a range from 3 to 11. These two exceptions tend to eliminate potential significant differences and the possibility of reproducible generalizations.

The first problem with the procedure was to allow the students the freedom to study the materials for three days. Apparently no data were gathered on the amount of time dedicated to this project by each student. Therefore, each student had the opportunity to work toward complete mastery of the 97 frames in the learning packets. Students with a weak background in the concepts or with poor skills of critical thinking could easily arrive at mastery of the assigned material by spending a relatively large amount of time in study. Some studies on individualized programmed materials have demonstrated that students will achieve mastery by using widely varying amounts of time, e.g., some students will use four or five times as many hours as other students. Probably all of the 121 subjects were capable of achieving mastery of the materials under the given circumstances. The most important variable may have been the motivation of the students toward this task. Some individuals strongly dislike self-instructional programs while others greatly enjoy them. The investigator concluded that "systematically written programmed instruction seems to enhance learning for some students." Yes, this study does show that the subjects learned by use of such materials. The study does not show any comparison of this type of learning with other methods and therefore provides little or no support for selecting systematically written programmed instruction as a method of instruction.

The investigators concluded that behavioral objectives meaningfully presented to some learners with a new cognitive task can enhance learning at the knowledge and comprehension levels. The data from this study limit the "some learners" to only two situations: (1) On PT_1 the group of 14 subjects with higher scores on the prior knowledge test (PKT) had significantly higher scores than the 17 subjects with lower scores on the PKT when these two small subgroups experienced Treatment 3. The higher scores on PT_1 were limited only to the knowledge level questions. These results were not replicated in PT_2. (2) On PT_2 the group of 17 subjects with higher scores on the PKT had significantly higher scores than the 13 subjects with lower scores on the PKT when these two small subgroups experienced Treatment 2. These results were limited to scores on the knowledge and on the comprehension questions and were not found on PT_1. Obviously the significant differences in this study are relatively rare events, based on small samples, and not replicating between the two posttests. The lack of consistent trends in the data seems to indicate that the significant differences may be generated by undetermined differences within the small samples. The abstractor has little confidence that the same two situations and only these two will produce significant differences if the study is repeated with another 121 students.

Bloom (1976) generalizes that half of the variation in achievement by students can be attributed to prior knowledge. This study does support the impact of prior knowledge by showing that some groups have improved scores on a posttest when they have higher scores on the prior knowledge test. Nevertheless, the significant differences are relatively rare and probably were eliminated in most comparisons by the uncontrolled factor of time spent on the learning. The investigator is correct in concluding that the identification of specific entering behaviors seems to be a significant variable for a learner's performance.

The investigator's conclusion that the domains of critical thinking ability and the levels of cognitive behavioral objectives presented in this study are independent of each other seems to be based on very weak evidence. As noted above, there are only two situations among the 16 similar comparisons which produced significant differences. The small samples in these two comparisons supported the well established recognition of the major impact of prior knowledge on future achievements in learning. The abstractor believes that there is a weakness in this study because it does not provide more evidence in support of impact from prior knowledge. The absence of significant differences on the basis of the scores from the Cornell Critical Thinking Test may be explained in terms of variations of time spent in studying the packets, variations in motivation toward self-instructional packets, and the size of the samples. The abstractor does not find the absence of significant differences to be sufficient basis for establishing the independence of critical thinking ability from the higher levels of the cognitive domain. Definitions from Bloom's Taxonomy (1956) seem clearly to require critical thinking ability in order to achieve the higher levels of cognitive thinking. The only conclusion which the abstractor could confidently support is the recognition that prior knowledge is the most significant variable in predicting achievement on self-instructional programs.

The investigators recognized that this study provides the implication for college science teachers to preassess students' prior knowledge. Certainly teachers will be most effective in their selection of curriculum when they clearly know the level of knowledge of students when they begin each unit of instruction. The investigators have been helpful to science educators by encouraging the use of preassessment to meet the unique needs of learners. In this study, preassessment was used to divide the subjects into groups of high and of low prior knowledge and into groups of high and of low critical thinking skills. In research studies, pretesting is usually given to determine the increments in achievement for each subject. The abstractor hypothesizes that if increments had been measured in this study, the subjects with low prior knowledge probably would have achieved significantly greater increments. This hypothesis is based on the finding of few significant differences among groups in the posttests. The abstractor's suggestion for another research study would contribute to the many studies already completed on self-instructional packets as a method of learning.

REFERENCES

Anderson, E. J.; H. T. DeMelo; M. Szabo; and G. Toth. "Behavioral Objectives, Science Processes, and Learning from Inquiry-Oriented Instructional Materials." Science Education 59(2): 263-271, 1975.

Ausubel, D. P. The Psychology of Meaningful Verbal Learning. New York: Grune and Stratton, 1963.

Bloom, B. S. (Ed.) Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain. New York: David McKay Co., 1956.

Bloom, B. S. Human Characteristics and School Learning. New York: McGraw-Hill, 1976.

Madaus, G. F.; E. M. Woods; and R. L. Nuttal. "A Causal Model Analysis of Bloom's Taxonomy." American Educational Research Journal 10: 253-262, 1973.

Hermann, G. D. and N. G. Hincksman. "Inductive versus Deductive Approaches in Teaching a Lesson in Chemistry." Journal of Research in Science Teaching, 15(1): 37-42, 1978.

Descriptor--*Achievement; *Chemistry; Deductive Methods; Educational Research; Inductive Methods; *Instruction; Science Education; Secondary Education; *Secondary School Science, *Teaching Methods

Expanded abstract and analysis prepared especially for I.S.E. by John R. Staver, De Paul University.

Purpose

The purposes of the investigators in this study are twofold. First, the efficacy of inductive and deductive teaching methods are compared in teaching a unit of chemistry. Second, trait-treatment interactions concerning subsets of learners and increased effectiveness of the inductive method are investigated.

Rationale

Proponents of inductive learning claim that student involvement in problem-solving activities aids retention and further discovery learning. However, a second claim often leveled by critics is that inductive problem solving activities require more time for students to reach identical learning goals than does a deductive approach.

Discovery learning experiments, according to the investigators, are generally concerned with sequence of instruction. In this study, inductive and deductive strategies are compared which are distinct in the sequence of instruction but are identical in time allotment. An inductive learning strategy is characterized by examples of a rule followed by a rule, whereas a rule followed by examples of the rule identifies a deductive method in this investigation.

Evidence in the literature of education and psychology indicates that differences in subject performance under inductive and deductive treatments may be expected with respect to IQ, trait anxiety, and sex. Specifically, high IQ subjects will perform inductive problem-solving activities better than average IQ students. High IQ children will learn equally well by an inductive or deductive strategy. Average IQ students will retain more information after inductive learning than after deductive learning. Further, highly anxious subjects will learn better deductively than inductively (Sakmyser, 1974). Finally, male students will outperform their female counterparts in the inductive learning of a chemistry task (Ormerod, 1975).

Research Design and Procedure

The independent variables in this factorial research design were instruction, trait anxiety, IQ and gender, whereas the dependent variables were the immediate and delayed posttests. The data were analyzed by analysis of variance.

Participants in the study were ninth grade advanced level science students from nine high schools in the Sydney, Australia, metropolitan area. There were 455 subjects involved, but 156 cases were not included in the data analysis because of absence, lack of IQ data, and clerical errors. Exclusion of these subjects left 299 cases (134 males and 165 females) who were randomly assigned to the inductive or deductive learning program on a within-class basis. The authors noted that subject wastage was random.

Deductive and inductive instruction was focused on stoichiometry, a unit in chemistry not yet encountered by the subjects. Learning materials included linear programmed instruction booklets and separate response sheets for each type of instruction. A fixed rate of instructional progress was assured by use of audiotapes. The lone distinction between the deductive and inductive booklets was in the relative placement of the rules to be taught. Each booklet contained 60 frames; the rule was placed in the eighth and tenth frames of a 10-frame set in the inductive format, whereas the rule was found in the first and third frames of a 10-frame set in the deductive format. A table of chemical symbols and valancies (authors' term) and an answer sheet were included in each program. All directions were administered via audiotape, and the time allotment (30 to 90 seconds per frame) was also controlled in this manner. The total learning period lasted approximately one hour. A posttest was administered immediately following instruction and a parallel form of the posttest was given two weeks later. A modified form of Sarason's Test Anxiety Scale for children was administered immediately following the second posttest. The investigators did not specifically mention the source of IQ data; it is a logical guess that such information was obtained from school records.

Findings

A summary of the authors' findings is given below:

1. The deductive instructional group scored significantly better ($p < .05$) on the immediate posttest than did the inductive instructional group, but no significant difference between treatment groups was revealed on the delayed posttest.
2. High IQ students did significantly better on the immediate and delayed posttests than did their average IQ counterparts.
3. No significant differences were observed on either posttest between the high and low test-anxious groups.
4. Females scored significantly better than did males on both posttests.
5. The interaction of method x IQ x anxiety reached statistical significance ($p < .05$) for the delayed retention test. All remaining interactions were not significant.

Interpretations

Conclusions, inferences, and implications made by the authors are summarized below:

1) Presence of a significant difference in favor of the deductive group on the immediate posttest and absence of such a difference on the delayed posttest are consistent with earlier findings.

2) A tentative suggestion, based on the findings, is that a deductive method may prove superior for immediate retention on difficult learning tasks, whereas the inductive method seems equally effective for delayed retention.

3) If discovery learning techniques are developed, time allotments are equal, and delayed retention is equally effective, then inductive learning is the advantageous method.

4. Absence of interactions among method, IQ, trait anxiety, and sex may have resulted from the high degree of structure associated with this learning task.

ABSTRACTOR'S ANALYSIS

The issue concerning the relative efficacy of inductive and deductive methods is the most important point of this investigation. The instance of equal effectiveness for delayed retention reported in this investigation is noteworthy because equal allotments of time were given to each instructional strategy. However, caution must be exercised in the results and generalization of these findings. The learning task was highly structured, brief, and involved the use of audiotapes. Also 33 percent of the students were not included in the analysis, due to missing data. Removal of one-third of the subjects from the analysis created a serious problem. The authors' note that the loss was random is difficult to accept without further evidence. The participants included in the data analysis may not be representative of the science class from which they were selected, thereby making further generalization tenuous. Techniques for handling unequal cell frequencies in ANOVA were available in the mid-1970's, and such methods are more appropriate than is exclusion of cases due to missing data. It seems best to view the results of this study with much caution. The findings have heuristic value, but further work is needed to support the evidence presented in this study.

REFERENCES

Ormerod, M.B. Pupil's Attitudes to Science. London: NFER, 1975.

Sakmyser, D. D. "Comparison of Inductive and Deductive Programmed Instruction on Chemical Equilibrium for High School Students." Journal of Research in Science Teaching, 14: 66-77, 1977

Krockover, Gerald and Marshall Malcolm. "The Effects of the Science Curriculum Improvement Study on a Child's Self-Concept." Journal of Research in Science Teaching 14(4): 295-299, 1977.

Descriptors--Educational Research; Elementary Education; *Elementary School Science; *Elementary School Students; Instruction; *Science Course Improvement Project; Science Education; *Self Concept

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

Purpose

To investigate an hypothesized relationship between participation in SCIS and the child's positive self-concept.

Rationale

Research studies have suggested that the learning environment, e.g., a humanistic emphasis, enhances the development of a strong positive self-concept in students. Other studies also support the conclusion that the stronger the self-concept of the student the greater the student's academic performance. Other variables such as gender, race and age that may enhance or detract from this relationship have also been investigated. The precise way in which a specific science curriculum as a context for learning influences the dependent variable of a child's self-concept is a key concern in this study.

Research Design and Procedure

The design was a pre-post control group where

$$O_1 \quad X_A \quad O_2$$

$$O_3 \quad X_B \quad O_4$$

O_{1-4} = measures of self concept with Pier Harris scale

X_A = SCIS curriculum for 4-1/2 months

X_B = conventional curriculum for 4-1/2 months.

The subjects used in this study were 189 students, third to sixth grades, in Indiana schools. The subjects were intact groups enrolled in eight classes. The experimental group consisted of two randomly assigned classes to receive the SCIS curriculum at each level, grades 3, 4, 5 and 6. The control classes continued to have a science textbook-based sequence as had been their custom.

All the subjects were given pretest of self-concept, Piers Harris Scale. The teachers were given the Bratt Attitude Test to ensure lack of bias in teacher attitude.

The treatment consisted of four and one-half months of instruction either in SCIS or in conventional textbook. At the end of the treatment the Piers Harris test was given again to all subjects.

Findings

Using a three-way (4 [grade] x 2 [gender] x 2 [treatment]) analysis of variance, the mean gain scores (total and sex cluster) from the Piers Harris Scale were analyzed. No evidence was found that gender or grade level produced a change in child's self-concept. The type of instruction was found to be a significant factor.

Interpretation

For this study, it can be concluded that the science curricula can be effectively used to enhance a child's positive self-concept.

ABSTRACTOR'S ANALYSIS

This study provides useful hints about the impact of school experiences on students' self-concept. That school experience would be expected to show such an impact is well supported by the authors. That school experience did indeed have this impact is far less certain. The design permits only the certainty that post-instructional measures of self-concept were indeed different than the pre-instructional measure of that variable. What is missing is a clear measure of the validity of the independent variable. In what ways was the instructional variable actually manipulated--and with what confidence can one believe that there was enough difference in the experimental and control treatment to attribute change in pre-post measure to the treatment? In the absence of either a specific description of the experimental treatment--and an indication that students had indeed acquired knowledge or skills, the reader is permitted only the conclusion that student performance on the measure did change over the four and one-half month period. This is reflected in the very limited implications for science teaching the authors were willing to make.

CRITIQUES

and

RESPONSES

Goodstein, M. and A. C. Howe. "The Use of Concrete Methods in Secondary Chemistry Instruction." Journal of Research in Science Teaching 15(5): 361-366, 1978.

Descriptors--*Chemistry; Cognitive Development; *Educational Research; *Learning Theories; Models; *Science Education; *Secondary Education; *Teaching Methods

Expanded abstract and analysis prepared especially for I.S.E. by John R. Staver, DePaul University.

Purpose

The purpose of the investigators in this study was to answer the following questions: First, will the use of concrete exemplars in chemistry instruction improve learning in chemistry? Second, is learning improvement related to the operational level of the student's thinking? The authors' stated hypothesis was that instructional methods in which learners use concrete models and exemplars of a concept will lead to better understanding by students at both the concrete and formal operational levels of cognitive development.

Rationale

The theoretical base of the investigation is founded in Piaget's theory of cognitive development. The instructional issue concerns the value of matching teaching strategies and the intellectual development of students for the improvement of student learning.

Research Design and Procedure

Ninety-five (95) chemistry students enrolled in a high school whose population is largely industrial lower-middle-class participated in the study. Each subject was a member of one of four regular chemistry classes; students in the honors chemistry section were excluded. The mean age of the subjects was 16 years, 8 months.

The four classes remained intact during the investigation; two were assigned to the treatment group such that the two teachers providing instruction each taught one experimental and one control class. The unit of instruction was stoichiometry, the first unit requiring chemical calculations for these subjects. Before instruction, the Piagetian level of each participant was ascertained. Next, stoichiometry instruction, control and experimental, was provided for six weeks. After instruction, student comprehension of stoichiometry was measured. A diagram of the research design appears in Figure 1.

Experimental Group (2 intact classes)	→	Piagetian Level Assessment	→	Stoichiometry instruction with concrete models and exemplars	→	Stoichiometry Posttest
Control Group (2 intact classes)	→	Piagetian Level Assessment	→	Stoichiometry instruction without concrete models and exemplars	→	Stoichiometry Posttest

Figure 1.--Design of the Study

Assessment of the participants' logical reasoning processes was done by a written instrument. The three sections of the written test measured ability to perform the operations of exclusion, combination, and proportional reasoning. Sections 1 and 2, exclusion and combination, were taken from Gray's (1973) validated test, whereas Section 3, proportional reasoning, was adapted from two tasks designed by Karplus (1975), "Mr. Tall-Mr. Short" and "the car problem."

The six-week instructional period was conducted in a lecture-response mode with associated drill and laboratory experiments. The stated principal difference between the treatment and control classes involved the use of concrete models (marshmallows, jujubees, styrofoam balls, and toothpicks) to represent atoms, molecules, and ions in stoichiometric chemical reactions. Laboratory work consisted of four two-hour periods during which all students performed the following experiments: 1) indirect count of large numbers of particles, 2) determination of the percent of water in a hydrate and 3) determination of the ratio of the product (NaCl) to reactant (Na_2CO_3). Control classes performed an experiment to determine Avogadro's number by monomolecular layer method, but treatment classes instead worked with concrete models and 0.1 mole samples of certain elements and compounds to develop the idea that the mass ratio of single atoms of different elements is the same as the mass ratio of models of the same element. All instruction in both groups was based on the introductory chemistry textbook by Metcalfe, Williams and Castka (1974).

Comprehension of stoichiometry was measured by two questions (shown below) embedded in the unit exam administered to all participants following instruction. The two questions were unlike any encountered by subjects in the text or during instruction. They were constructed to assess the development of a conceptual understanding of stoichiometry in contrast with the use of a memorized dimensional analysis algorithm in the solution of quantitative problems. All other items in the unit test were similar to instructional examples. The chi-square statistic was employed to test the null hypothesis.

Question 1: Which, if either, has more atoms, 30 grams of oxygen or 30 grams of chlorine? Explain your answer.

Question 2: 35 grams of chlorine are reacted with 35 grams of sodium to form NaCl. Is there exactly enough of each or will sodium be left over or will chlorine be left over? Explain your answer.

Findings

A summary of the authors' stated findings is provided below:

1. The expected wide variation in cognitive level of the participants was observed (C1--early concrete: 4 percent; C2--advanced concrete: 21 percent; F1--early formal: 35 percent; F2--advanced formal: 40 percent).
2. Participants' scores on the stoichiometry comprehension posttest were significantly associated ($X^2 = 25.68$, $df = 4$, $p < .0001$) with cognitive level within the treatment group, but posttest scores

were independent ($X^2 = 8.94$, df = 4, N.S., $p < .05$) for subjects in the control group.

3. For subjects classified as advanced formal operational, a significant number in the treatment group achieved higher scores on the posttest than their counterparts in the control group ($X^2 = 7.36$, df = 2, $p < .05$). The independence hypothesis was not rejected for students classified as early formal or concrete.

Interpretations

The conclusions, inferences, and implications made by the authors are summarized below:

1. Concrete operational students did not profit from instruction including concrete models and exemplars.
2. The results are unclear concerning students classified as early formal operational.
3. Instruction with concrete models and exemplars was beneficial for advanced formal thinkers.
4. Negative findings concerning the benefits of the instructional treatment for concrete learners cannot be attributed to poorly designed procedures.
5. The findings do not support the view that any subject may be successfully taught if concrete methods are used.
6. Three factors (cognitive level of the learner, conceptual level of the material, and the method of instruction) must be considered in the interpretation of the results.
7. The concepts were apparently too difficult for all but the advanced formal thinkers.
8. The instructional method was a controlling factor for advanced formal students. Concrete exemplars provided a base for formal thought by these participants in the experimental group, but such thinking was apparently not used by advanced formal control subjects who did not have such models.

9. The findings suggest that concrete thinkers cannot learn concepts that require advanced formal thought regardless of the instructional method, but the use of concrete models and exemplars by formal thinkers can enhance learning of formal concepts.

10. The first consideration should be level of thought required in the concepts learned, then a reasoning level assessment of the learners. If the students do not yet possess the necessary cognitive reasoning patterns, then instruction should be postponed or the concepts reduced in abstractness until they are consistent with nonformal reasoning patterns.

ABSTRACTOR'S ANALYSIS

The authors have provided evidence concerning a most important instructional question: Will matching the conceptual level of instruction and the intellectual level of students improve learning? Whereas generalizations beyond the chemistry student population in the school are unwarranted (and the authors rightfully make none) due to the nature of participant selection and assignment to groups, the findings, conclusions and interpretations merit discussion within the limitations of the study.

Three points of concern require deliberation. First, the planned instructional difference between the treatment and control groups was the use of the aforementioned concrete models. However, control group participants were required to do an experimental determination of Avogrado's number by monomolecular layer method, whereas the experimental group performed an alternate experiment. Further, the investigators noted that Avogrado's number experiment required a high degree of abstract (formal) thought, but the alternate experiment seemed inherently less abstract. This represents a very substantial difference; it involves 25 percent of the time devoted to laboratory work during instruction.

The second concern stems from the lack of detail in the description of the treatment. Readers are merely informed of the planned instructional difference involving the use of concrete exemplars. A more detailed summary of the frequency and methods with which concrete models were used during the six-week instructional period is needed, especially in view of the findings. Without knowledge of such details, readers must question the degree of any difference between instruction in the experimental and control groups. Leonard and Lowery (1978) have empirically addressed the issue concerning differences in instructional treatments, and similar procedures should be an integral part of empirical effectiveness comparisons. The results, however, do reveal that some instructional differences existed between the groups, but the magnitude of that difference cannot be inferred from the authors' cursory description of the instruction in the treatment and control groups. The central point of the second concern is this: Can the expected comprehension by concrete thinkers in the treatment group be better than that of their control group counterparts, given this particular instructional treatment and posttest? In my view, the answer is no, until the treatment is detailed. An explanation of my answer includes an integration of the authors' posttest, the treatment, and the nature of cognitive development according to Piaget. The posttest is genuinely a test of comprehension, or possibly application (Bloom, 1956), and such cognitive thought implies a deeper internalization of concepts than a commitment to memory. The investigators are to be applauded for setting such cognitive goals as their learning objectives. However, a problem arises with respect to the instruction and concrete thinkers. Possibly, an insufficient frequency and integration of the concrete models and exemplars did not allow the concrete thinkers to internalize stoichiometric concepts to the degree measured by the posttest. It seems clear that for formal thinkers, the frequency and integration was quite sufficient, and the concrete models did, as the authors state, provide a nonsymbolic framework for carrying out the formal reasoning patterns already within their capability. But a more frequent use and elaborate degree of integration must occur for the concrete thinker to employ concrete reasoning patterns in the solution of such problems, especially without the use of models.

The third concern is derived from the mental capabilities of concrete thinkers. Again, in my view, the treatment possibly did not provide enough integration and reduction of abstractness within the time frame of the experiment for concrete thinkers to be successful on such posttest problems. This study was not a training study, and the authors rightfully did not expect concrete thinkers to begin using formal reasoning patterns as a result of treatment. But, was the treatment effective in helping concrete thinkers understand stoichiometry via concrete reasoning patterns to the extent that such participants could solve posttest problems at Bloom's (1956) comprehension or application level without concrete models? It would have been most interesting to inspect the posttest results of concrete thinkers who used the models during the posttest. The abstract structural relationships so apparent to formal thinkers are not obvious to concrete thinkers. If the concrete aids are removed, the concrete thinker is left to consider the problem abstractly. Whereas the treatment was almost certainly helpful, it probably did not adequately prepare such subjects for the cognitive rigors of posttest.

The entire line of argument may be due, quite possibly, to factors beyond the control of the investigators. Herron (1977) has often taken issue with the desires of reviewers and editors for brevity to the extent that clarity is comprised. All concerned must remember that their goal is to communicate as much as possible. In this case, the reader is left with questions due to a lack of clarity concerning the instruction.

The opportunity to consider the empirical data in this report was most pleasing. However, a detailed inspection revealed an inappropriate analysis of part of the data. The point involves minimum acceptable expected cell frequencies in a chi-square contingency table. Siegel (1956) notes that when the number of columns in the contingency table is larger than 2, the X^2 test may be used, provided that no more than 20 percent of the cells have an expected frequency of less than 5, and no cell has an expected frequency of less than 1. A X^2 test is not meaningful unless these criteria are met. Unfortunately, the expected frequency falls below 5 in one-third of the cells for each X^2 done. Siegel (1956) recommends the combination of adjacent categories to alleviate the problem. Whereas combination of adjacent categories represents an acceptable statistical procedure, it should not be arbitrarily done. In this investigation, the adjacent categories of "early formal" and "advanced formal" thinkers also represent a theoretically (psychologically) acceptable combination. The categories have more in common than either does with the "concrete" classification.

When the X^2 tests performed by the authors are done on the modified contingency table, the following results are obtained:

1. X^2 tests remain meaningless for comparisons of cognitive level and score within the experimental and control groups due to the aforementioned criteria concerning minimum expected cell frequencies.

2. For all formal thinkers, the proportion of students in the experimental group who obtained high scores is significantly greater ($X^2 = 12.92$, $df = 2$, $p < .01$) than the proportion of control group subjects.

In sum, the contribution of this investigation, in its present form, to the science education literature remains unclear. The exact nature of the use of concrete models and exemplars needs more detailed description, and data require analysis by more appropriate methods. The score distribution could be collapsed further by combination of adjacent categories to give a 2 x 2 contingency table. The Fisher Exact Probability Test (Siegel, 1956) is an appropriate alternative to X^2 for 2 x 2 contingency tables with small expected frequencies. When the above suggestions or other acceptable alternatives are carried out, the findings, conclusions, and implications will be clarified and the contribution of this investigation will become more lucid.

REFERENCES

Bloom, B. S. (Ed.). Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook 1: Cognitive Domain. New York: David McKay Co., 1956.

Gray, W. M. Test of Logical Thinking. Dayton, Ohio: Dayton University, 1973. (ERIC Document Reproduction Service No. ED 109 197)

Herron, J. D. "Expanded Abstract and Analysis of: 'Critical Thinking Ability, Openmindedness, and Knowledge of the Processes of Science of Chemistry and Non-Chemistry Students'." In S. Helgeson and P. Blosser (Eds.), Investigations in Science Education 3(2): 31-37, 1977.

Karplus, R.; E. Karplus; M. Formisano; and A. Paulson. Proportional Reasoning and Control of Variables in Seven Countries. Berkeley, California: Lawrence Hall of Science, University of California, 1975.

Leonard, W. H. and L. F. Lowery. "Was There Really an Experiment?" Educational Researcher 6: 4-7, 1979.

Metcalfe, H. C.; J. E. Williams; and J. F. Castka. Modern Chemistry. New York: Holt, Rinehart and Winston, 1974.

Siegel, S. Nonparametric Statistics. New York: McGraw-Hill, 1956.

IN RESPONSE TO THE ANALYSIS OF

Goodstein, M. P. and Ann Howe. "The Use of Concrete Methods in Secondary Chemistry Instruction," by John R. Staver. Investigations in Science Education, 8(2): 37-45, 1982.

by

Madeline P. Goodstein
Central Connecticut State College
and
Ann Howe
Syracuse University

We would like to respond to the three points of concern stated by Professor Staver as follows:

On the Difference in Instructional Treatment

It is essential to this type of study that both the control and experimental groups receive identical time periods of instruction. This means that the time spent on the special instructional treatment with the experimental group must be equalled by time spent with the control group on some different treatment of the same topic. In this case, the use of hands-on models required time in the experimental treatment that was devoted to such nonconcrete activities as oral discussion or working out problems with the control group.

In the same vein, some laboratory time also was used to increase the amount of concrete activity available to the experimental group. Of the four experiments carried out, only one, the monomolecular layer experiments, was judged to be essentially formal rather than concrete. The hands-on activity in this experiment is not in itself productive of understanding of the size of a molecule; extensive mathematical and geometrical instructions are required to reach the desired conclusion. Hence, an experiment which used a more concrete procedure was selected to use the same time slot as was filled by the monomolecular layer experiment for the control group. This time period accounted for seven percent of the total course time.

On Loss of Detail in the Description of the Instructional Treatment

As part of the treatment selected for the experimental group, models were frequently picked up and shown to the class for rather short periods and then put down again. Also, the models to be used in any discussion were visible on the front desk throughout that period; some models were posted permanently on the wall. We question whether any attempt to modify this would be useful.

The editor's limitation on the length of the article unfortunately forced us to eliminate many of the details of the instructional procedures. However, there are only a specific number of places in the topic of stoichiometry which can be concretized with models. We believe that similar use of the models would be made by any teacher offering secondary instruction in stoichiometry who sought to replicate this study.

On the Chi Square Analysis

A more serious point is the reviewer's statement that our use of chi square for the analysis of the data was inappropriate. We cannot accept this. McNemar (1962) discusses the effect of small E's on the value of chi square and states that "the effect of small E's in producing discontinuities is not as marked when df is 2 or more" and "there is evidence that, when df is not small, E's as low as 2 will not produce misleading chi square values" (p. 218). We would, of course, have felt more confident of the results of the chi square test if we had had fewer cells with low frequencies but we do not think the reported analysis is misleading. By publishing the frequencies in a clear and unequivocal style we made it possible for the reader to draw his or her own conclusions about our use of statistics and interpretation of results. The reviewer's suggestion that categories be collapsed does not seem to solve the problem since that would obscure the distinctions between students at the three cognitive levels.

In our judgment we chose the best available alternative and we do not think the results are meaningless.

It should be remembered that we did not claim to have conducted a definitive experiment--it was, instead, an effort to test an idea on a small scale under conditions that exist in real classrooms. We would welcome evidence from others who have been able to improve on our methods.

REFERENCE

McNemar, Q. Psychological Statistics. New York: John Wiley & Sons, 1962.

Quorn, Kerry Charles and Larry Dean Yore. "Comparison Studies of Reading Readiness Skills Acquisition by Different Methods: Formal Reading Readiness Program, Informal Reading Readiness Program, and a Kindergarten Science Program." Science Education 62(4): 459-465, 1978.
Descriptors--Comprehension; Elementary Education; *Instruction; *Kindergarten; *Language Ability; *Perception; *Process Education; Reading Achievement; *Reading Readiness; Science Education

Expanded abstract and analysis prepared especially for I.S.E. by Donald E. Riechard, Emory University.

Purpose

The primary paper reports on two different but closely related experiments. The purpose of experiment one was to assess the effectiveness of two dissimilar programs, Science--A Process Approach (S--APA) and The First Talking Alphabet (FTA), upon the acquisition of reading readiness skills of kindergartners. The hypotheses tested were: There will be no significant

1. treatment, sex, and attendance-time main effects;
2. interaction effects;
3. differences between pretest and posttest means on any of the reading readiness measures.

The purpose of experiment two was to compare the effects of four different programs (S--APA, FTA, informal language development, and a control). The hypotheses tested were: There will be no significant

1. treatment, sex, or teacher main-effects;
2. interaction effects on any of the reading readiness measures.

Rationale

The authors build a rationale on published research. References are cited to support the following:

1. There are predictive correlations of certain skills with later reading achievement. (Skills are listed in the primary paper.)

2. The skills, above, may be developed through training.

3. Similarities between new kindergarten science curricula and teachers' informal readiness activities suggest that science processes and reading readiness skills are not mutually exclusive.

4. Certain science activities might provide an opportunity and climate for reading readiness skills acquisition.

Research Design and Procedure

As stated above, two different but closely related experiments were presented by the authors. Designs and procedures of the experiments are given under separate subheadings below.

Experiment I. Fifty-one pupils from morning and afternoon kindergarten classes were randomly assigned to four treatment groups. Two reading readiness measures, Metropolitan Readiness Test and Clymer-Barrett Prereading Battery, were administered on a pre- and posttest schedule. One morning group and one afternoon group were randomly selected for one of the treatment programs (S--APA or FTA). The two remaining groups were assigned the other program. Treatment consisted of an investigator teaching the two instructional programs (S--APA and FTA) over a ten-week period. Each group received 22 half-hour periods of instruction.

Upon completion of the pretest, treatment, and posttest, the reading readiness measures were scored. Posttest data were analyzed by a three-way analysis of variance, using treatment, sex, and attendance time (morning or afternoon) as the main dimensions. Pretest-posttest mean gain scores for each treatment, sex, and attendance-time group were analyzed by correlated t-tests.

Experiment II. Fifty-four kindergarten pupils from one school were randomly assigned to three treatment groups and one control group. The treatment groups were science, formal reading readiness, and informal reading readiness. One of the investigators conducted 14 sessions with each of the treatment groups. The science sessions were lessons from

S--APA while the formal reading readiness lessons were selected from FTA. The informal reading readiness sessions were developed by one of the authors. The control group continued with the regular kindergarten program.

A posttest was administered after treatment. The posttest consisted of three measures of reading readiness--the Clymer-Barrett Prereading Battery, the Wepman Auditory Discrimination Test, and the Moe Beginning Consonant Sound Symbol Relationship Test.

Analysis of data was done by means of a three-way analysis of variance with treatment, sex, and teacher as the independent variables. The teacher variable was defined as the two "regular" kindergarten teachers. Reading readiness was the dependent variable.

Findings

Experiment I. There were significant ($p \leq 0.05$) differences in the following:

1. Attendance times on the Clymer-Barrett discrimination of beginning sounds subtest.
2. Treatment-by-attendance time interaction on the Metropolitan matching subtest.
3. Treatment-by-sex-by-attendance time interaction on the Metropolitan copy subtest.

The t-test analysis of gains between pretest and posttest means yielded 51 significant ($p \leq 0.05$) findings and 13 at or greater than the 0.10 level of probability. These data are given by table in the primary article.

Experiment II. Results may be summarized as follows:

1. Significant ($p \leq 0.05$) teacher effects were found for the auditory discrimination section of the Clymer-Barrett.

2. A significant ($p \leq 0.05$) sex effect was found for the Wepman Test of Auditory Discrimination.

3. No significant treatment differences or interaction effects were found.

Interpretations

The authors summarize their research by the following:

1. The effectiveness of kindergarten science activities on reading readiness development cannot be supported conclusively by these studies.

2. The theoretical links between reading and the science processes --observing, inferring, and classifying--were not evident in the empirical data.

3. Kindergarten science activities effectively develop science skills without interfering with reading readiness attainment. Thus, science activities may be reasonably included in the curriculum even though their effectiveness for reading readiness development has not been demonstrated.

ABSTRACTOR'S ANALYSIS

Contributions of the Study

The authors' summary of findings (above) raises questions about the contributions of this study to existing knowledge on relationships between reading readiness and science experiences. Is this a unique case of investigators being a bit overly conservative in interpreting their data?

At any rate, the study does provide useful insight into methodology for research with young children. Further, it suggests some lessons on interpreting data and preparing written reports.

Research Design and Validity

The pretest-treatment-posttest design used in experiment one seems adequate for the specific purposes stated. However, the report does not indicate if a comparison of pretest scores of the different groups was made. Random assignment to treatments was used and each treatment contained a morning and afternoon group. Still, it would be helpful in analyzing results to know how the groups compared prior to treatment.

Experiment two was conducted with treatment-posttests only; no pretests were administered. Again, despite randomization, the question of group comparisons before treatment is raised.

Several other questions are related to design and procedure. Among them are the following:

1. Was the same school used in experiment one and experiment two? If so, were any of the same children used in both experiments? Were the same "regular" teachers involved in both experiments?
2. In experiment one, were morning and afternoon classes taught by the same "regular" teacher, or were different "regular" teachers involved? The teacher variable was significant on some measures in experiment two.
3. Did experiment two use both morning and afternoon groups? Attendance time (morning or afternoon) was significant on some measures in experiment one.
4. How long did experiment two last? Experiment one lasted 10 weeks. The authors suggest that experiments of this nature which are concluded in a short period of time are less likely to produce significant results than experiments running an entire year.

Research in Early Childhood

The design of the study indicates a sensitivity to some of the special problems encountered in early childhood research. Good efforts were made at randomization and assignment to treatment and attendance-time groups. Also, treatment and testing were rotated across groups. Such rotation can minimize the influence of extraneous variables such as children's attention spans, fatigue, eating habits, napping and play patterns, and so on.

The paper does not indicate if children were familiar with the investigators but this is also a factor to be considered. It is sometimes advisable for a stranger to make several "rapport visits" before actually beginning an experimental treatment and/or testing with young children.

Interpreting Results

The authors were very conservative in drawing conclusions and interpreting results. In general, brevity and caution in interpreting behavioral research are highly commendable. Care must be taken, however, not to overlook relationships or inadequately analyze data for the sake of concise and cautious reporting.

Quite appropriately, the investigators do not overgeneralize or make statements not fully supported by the data. They even question some of their statistically significant findings and suggest that the results be "viewed with suspicion."

Another example of caution is in the summary statement that "kindergarten science activities effectiveness on reading readiness cannot be supported conclusively by these studies. The theoretical links between reading and the science processes...were not evident in the empirical data." These conclusions are made despite the large number of significant pretest to posttest gains in reading performance by the S--APA groups in experiment one. Since similar gains were made by the FTA groups, the authors allow only that both programs produced similar types of achievement and that further research is needed to isolate the significant variables.

Written Report

At times it was difficult for this abstractor to follow the two different experiments in one article. Keeping track of purposes, treatments, teachers, groups, results, et cetera, for each experiment was bothersome. Publication of two separate reports would have made easier reading and might have permitted a more thorough presentation of data and analysis of results.

Null hypotheses were stated for each of the experiments. However, in the interpretation of data, no mention was made of the hypotheses. Direct reference to previously listed hypotheses and a statement about the acceptance or rejection of each can be an excellent starting place for discussion of results. The writer of the report then has something specific to focus upon as conclusions and interpretations are made. Such a focus reduces the chances of <u>over-interpretation</u> of data, and, conversely, reduces the chances of <u>under-interpretation</u> (too cautious an interpretation) as well.

IN RESPONSE TO THE ANALYSIS OF

Quorn, Kerry C. and Larry D. Yore. "Comparison Studies of Reading Readiness Skills Acquisition by Different Methods: Formal Reading Readiness Program, Informal Reading Readiness Program, and a Kindergraten Science Program" by Donald E. Riechard. Investigations in Science Education, 8(2): 48-54, 1982.

by

Larry D. Yore
University of Victoria

The opening comments regarding the contributions and conservative interpretation of the data appear to be unsubstantiated by the majority of the related criticisms. Surely two independent studies with rather consistent findings and reasonable design and limitations make some contributions "to existing knowledge on relationships between reading readiness and science experiences." The critical question appears to be whether science educators can make unconditional claims regarding science experience's influence on reading achievement. Several researchers have found significant correlations and differences favoring activity science (Ayers and Mason, 1969, Kellogg, 1971; Morgan, Rachelson and Lloyd, 1977; Esler and Midgett, 1978), but others have not found such results (Ritz and Raven, 1970). Couple these inconsistent findings with the likelihood the nonsignificant results are less likely to be published and one sees a cloudy picture of the reading-science question. As a science educator I would have been most pleased to publish results that enhance the position of science in the elementary school curriculum. However, my research results did not support such an interpretation.

The issue concerning "conservative interpretation" must be viewed in terms of statistical expectation regarding rejection levels and the contribution of chance in finding significant differences (Hays, 1963, pp. 167-171). In Experiment I three-way analyses of variance were run of 14 measures of reading readiness, i.e., treatment, sex, attendance time, treatment x sex, treatment x attendance time, sex x attendance time, treatment x sex x attendance time. Therefore, on each measure of the dependent variable seven different hypotheses were tested, resulting in a total of 98 tests ($7 \times 14 = 98$). With a rejection level of 5 percent ($\alpha \leq 0.50$), one would expect five significant differences in 100 tests purely due to chance. Therefore, when testing 98 hypotheses one should not be surprised to get 1, 2, 3, 4, or even 5 significant differences. On the other hand, if a small number of significant differences outline a rather consistent pattern, then greater concern should be expressed. Such a consistent pattern was outlined by the teacher factor in Experiment II. Three-way analyses of variance were run on 12 measures of the dependent variable, thus 84 tests were run ($7 \times 12 = 84$) with a rejection level of 5 percent ($\alpha \leq 0.05$). The expectation was that four significant differences would be found due to chance. The analysis of data in Experiment II yielded four significant differences but three of these involved the teacher factor favoring a more structured approach.

Design and Validity

The concern regarding the analysis of pretest data to assess equivalence of the random samples in Experiment I was a noticeable weakness in reporting. A complete analysis of pretests were completed and reported in the more complete document (Yore, 1973, pp. 55-61). The analysis of variance on pretest data yielded two significant (p . 0.05) differences on the 98 hypotheses tested. Only the attendance time x treatment x sex interactions for Clymer-Barrett Beginning sounds and Word Matching were found to be significantly (p 0.05) different.

Since random samples were used, differences that existed were attributed to chance and the random sampling fulfilled the assumption of the ANOVA. On the other hand, since Experiment I used a pre-post test design, the descriptive statistics or summary of results should have been provided for the reader. Experiment II attempted to avoid the pretest-treatment interaction and test-retest effect; therefore, no analysis of pretest was possible and is not a legitimate criticism since the pretest was removed to increase the external validity of the design (Campbell and Stanley, 1963).

Concern regarding the school populations are implicitly answered with the mention of "kindergarten class(es) of a Victoria, British Columbia, Canada, public school" (Quorn and Yore, p. 460) for Experiment I and "kindergarten pupils (54) from a school near Victoria, British Columbia" (Quorn and Yore, p. 462) for Experiment II. The key operants "of" and "near" indicate that different schools were used, therefore different teachers and pupils.

The pupils in Experiment I were taught by the same teacher, which allows for a three-way analysis of variance rather than four-way. In Experiment II two kindergarten classes which met simultaneously with different teachers were used as the sample. The simultaneous attendance allowed the disregard of attendance time and the addition of a teacher factor in the three-way analysis of variance. In both experiments subjects were randomly assigned to groups and the groups were randomly assigned a treatment.

The investigator conducted 14 instructional sessions with the assigned treatments over a 12-week duration in Experiment II. In Experiment I a four-week physical education (tumbling) unit was conducted by the investigator prior to the pretesting to establish rapport with the students. In Experiment II the investigator visited the classes regularly for a month prior to the start of the treatments.

Data Analysis and Interpreting Results

The comments that "care must be taken, however, not to overlook relationships or inadequately analyze data for the sake of concise and cautious reporting" is viewed with interest. The choice of statistical analysis was built into the design of both experiments a priori and the underlying assumption of the specific statistics were considered. Alternative statistical technique could have been used, such as in Experiment I--an analysis of covariance (Campbell and Stanley, 1963). The statistics were chosen on the advice of statistical consultants at three different universities (Victoria, Minnesota, and Washington State).

The concern regarding conservative interpretation is a criticism that may be taken by many as a positive attribute. The large number of significant pre-post test gains suggest that learning occurred over the duration of Experiment I. The lack of significant differences between treatments prevented any specific identification of the possible cause. The design of Experiment I did not allow the investigators to partition the variance down to resultant components. It was believed that the test-retest effect, the regular kindergarten program and outside influences contributed so much variance that the variances due to the specific treatments were inconsequential. The design of Experiment II attempted to factor out some of the effects. Unfortunately the regular program was not significantly different from the other three treatments. It was judged responsible not to interpret too liberally since the data and analyses appear rather consistent.

Given this opportunity to expand on the original interpretation, it appears that a potentially strong outside influence that might have discounted the effect of the treatments is television available in Victoria, British Columbia, Canada. Victoria is located near the USA-Canada border and has 12 channels of US and Canadian TV available for viewing. Several channels carry educational programs for young children, i.e., CBC, CTV, PBS and the US national networks. Reading readiness programs like "Sesame Street" viewed several times might be discounting the effect of formal in-school readiness programs.

REFERENCES

Ayers, J. and G. Mason. "Differential Effects of Science: A Process Approach upon Change in Metropolitan Readiness Test Scores Among Kindergarten Children." The Reading Teacher, 22: 435-439, 1969.

Campbell, D. and J. Stanley. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand-McNally College Publishing Company, 1963.

Esler, W. and B. Midgett. "Using Laboratory Experiences to Teach Reading." Science and Children, 33-34, February 15, 1978.

Hays, W. Statistics for Psychologists. New York: Holt, Rinehart and Winston, 1963.

Kellogg, D. "An Investigation of the Effect of the Science Curriculum Study's First Year Unit, Material Objects, on Gain in Reading Readiness." Doctoral dissertation, University of Oklahoma, 1971. University Microfilms. No. 71-27, 623.

Morgan, A.; S. Rachelson; and B. Lloyd. "Sciencing Activities as Contributors to the Development of Reading Skills in First Grade Students." Science Education, 61: 135-144, 1977.

Ritz, W. and R. Raven. "Some Effects of Structured Science and Visual Perception Instruction among Kindergarten Children." Journal of Research in Science Teaching, 7: 179-186, 1970.

Yore, L. "A Comparison Study of Reading Skills Acquisition by Two Methods: A Traditional Reading Readiness Program and a Kindergarten Science Curriculum." Unpublished Ph.D. dissertation, University of Minnesota, Minneapolis, 1973.

Cavin, Claudia S. and J. J. Lagowski. "Effects of Computer Simulated or Laboratory Experiments and Student Aptitude on Achievement and Time in a College General Chemistry Laboratory Course." Journal of Research in Science Teaching 15(6): 455-463, 1978.

Descriptors--*Achievement; *Aptitude; *Chemistry; College Science; *Computers Assisted Instruction; Educational Research; Higher Education; Individualized Instruction; Science Education; *Simulation

Expanded abstract and analysis prepared especially for I.S.E. by Joseph C. Cotham, Indiana University-Bloomington.

Purpose

This study investigated the effectiveness of computer simulated laboratory experiences in a college chemistry laboratory course. The following null hypotheses were tested:

1. There is no significant difference in achievement or time to complete the experiment of students doing a simulated or laboratory experiment.
2. There is no significant difference in achievement or time to complete the experiment and calculations of students of different aptitude.
3. There is no significant interaction between experiment type and student aptitude.

Rationale

The authors cite current interest in computer-assisted instruction (CAI) and computer simulated experiments (CSE) as possible means of improving student learning and providing for individual differences. Recent interest in microcomputer applications in the schools and the increasing accessibility of these systems does, indeed, emphasize the importance of generating a sound empirical basis for decisions concerning the use of computer systems in instruction. Evidence exists to support the use of CSE in place of laboratory experiments in high school, the use of CAI to teach the operation of an instrument, and the use of

CAI in supplementing college chemistry coursework. This study attempted to extend understanding of the utility of CSE in college chemistry laboratory instruction. An important aspect of this understanding, emphasized in the authors' rationale, is the relationship between aptitude and teaching method.

Research Design and Procedure

The hypotheses of this study were tested using a 2 x 2 factorial design (posttest only or time). The factors were method of doing the experiment (simulation or laboratory) and student aptitude (high or low scores on the SAT). The dependent variables were scores on written achievement tests, the time required to do the experiment, time required to perform experimental calculations, and in one test, achievement (score and time) on a performance test. The achievement tests were multiple choice instruments of varying lengths. The content validity of these tests was estimated informally using consultation with subject-matter experts. Reliability was estimated with the KR20 and Spearman-Brown formulas (values for the 10 tests ranged from .61 to .89). The performance test, which was administered individually, was based on a checklist of observable behaviors required to make a spectrophotometer measurement on a solution.

The sample consisted of students enrolled in an introductory chemistry course at The University of Texas in the fall of 1976. The course consisted of multiple laboratory sections which were offered at four different time periods. Pairs of laboratory/simulation experiments for each course topic (refer to Table I) were prepared and randomly assigned to each time period. Within each time period laboratory sections were randomly assigned to either a laboratory or simulation experiment. The number of different sections involved in each treatment was not reported. Absence of bias in the registration of students in one section or another within a time period was assumed and used to justify failure to randomly assign students to treatments. The treatments, which were only briefly described, appear to differ significantly in ways not attributable to type of experiment. For example, the simulated experiment for Group 1 involved a spectrophotometer, but the laboratory experiment did not. Also, treatments were administered in different laboratory sections. It was not clear from the report what steps were taken to deal with the threat to internal validity implied by this arrangement.

TABLE I

Sequence of Course Topics

Group	Topic
1	Kinetics
2	Wavelength of maximum absorbance, Beer's law, and determination of copper concentration
3	Element identification using emission spectroscopy
4	Beer's law and equilibrium constant determination

Dependent measures were administered at the conclusion of each experiment. Results were analyzed using a two-way analysis of variance.

Findings

The following findings were reported:

1. Based on achievement on written tests, students who did the simulation in Group 4 performed significantly better than students who did the laboratory experiment. Also, students who did the simulation in Group 2 performed significantly better on one dependent measure than students who did the laboratory experiment. No significant differences between experiment types were found for scores on the performance test taken by Group 2.

2. Students who did the simulated experiments in Groups 1 and 4 used significantly less time to perform the experiments than did students in the laboratory. However, students who did the laboratory experiments in Groups 1 and 3 used significantly less time for the calculations.

3. Students in the higher aptitude groups performed significantly better on all written tests than students in the lower aptitude group.

4. Even though no significant aptitude by treatment interactions were found (at the 0.05 level), evidence of ordinal interaction was obtained from one test from both Group 1 and Group 2 (greater increase for low aptitude group). Evidence of disordinal interaction was obtained from values of time to do the experiment for Group 4 (less time for low aptitude group in laboratory treatment).

Interpretations

The authors consider their study to be one experiment with three replications. Consequently, on the basis of significantly better performances in some of the experiments, they conclude that the use of CSE in place of laboratory experiments in college chemistry laboratory is supported. The conclusion that higher-aptitude students perform significantly better on all written tests is consistent with previous studies. Lack of significant differences on the performance test which required spectrophotometer manipulation was used to support the assertion that CSE can be used to teach the use of some types of laboratory instruments. The authors provided a detailed explanation of the observed differences in time to perform the experiment and time to perform experimental calculations. They speculated that the simulation subgroups of Groups 1 and 3 required more calculation time because the calculations they were required to perform were more complicated than the calculations done by the laboratory subgroups. The interaction results, although only suggestive, agree with previous results at the elementary level (Martin, 1973), indicating that CSE may be especially useful for low-aptitude students.

ABSTRACTOR'S ANALYSIS

The laboratory, possibly the hallmark of science education, deserves the critical scrutiny exemplified by this study of the effect of teaching method and aptitude on achievement in a college chemistry lab course. It is worthwhile, as an introduction to the analysis of this study, to refer to a landmark study of laboratory teaching conducted by Yager, Englen, and Snider (1969). On the basis of this study of different methods of teaching an adapted version of the BSCS Blue version to eighth graders, it was concluded that the laboratory approach provided no measurable advantages over other modes of instruction except in the development of laboratory skills. Cavin and Lagowski, in their study of college laboratory instruction, conclude that on all measures except time CSE accomplishes as much (if not more) than laboratory experiments--even in the development of laboratory skills! Are we then to conclude that the advent of CSE has sounded the death knell of laboratory experiences? Before any such conclusion can be made, it is necessary to examine some of the characteristics of this study more closely.

A notorious problem in studies of the effects of particular teaching methods is failure to adequately describe the method studied. The study addressed by this abstract suffers from this deficiency. What distinguishes a CSE experiment from a laboratory experiment? The simulated experiments for Groups 1, 2 and 4 involved manipulation of a spectrophotometer. If both CSE and laboratory experiments involve manipulation of laboratory instruments, what are the distinctive features of CSE that commend it to our use? The authors, to their credit, cite the importance of attending to the type of chemistry experiment that is simulated. The experiments in their study were conducive to simulation because they involved use of instruments. And yet, not all aspects of instrument use were simulated because instruments were used in some simulated experiments. A fuller description of the study's treatments would have dispelled some of these ambiguities.

Another difficulty in interpreting the results of this study are the potential threats to internal validity that were referred to earlier. The existence of multiple laboratory sections implies that treatments were administered by different instructors. Was anything done to counter the confounding effects of the instructor on treatments? And what steps were taken to insure that different administrations of the same treatment were uniform? This is especially salient for the sections receiving the laboratory experiments because of the presumed instructor interaction with this treatment. Failure to randomly assign students to treatments, although understandable under the naturalistic constraints of the experiment, deserves more comment. The assumption of lack of bias in distribution of students among sections may be valid. But situations exist where students in a particular program (e.g., elementary education majors) end up in one or only a few sections of a course due to scheduling requirements. Was this possibility investigated? If so, its discussion would have assisted in interpreting the study.

A number of tests were used in this study (10 written tests and 1 performance test). However, their focus appears limited to cognitive and psychomotor domains. It may have been an interesting addition to this study to assess students' attitudes toward chemistry and teaching method. These findings could provide useful insight into the utility of CSE.

The authors of this study devote considerable attention to the dependent variable, time: both time to perform the experiments and time to do experimental calculations. This is appropriate because the laboratory has been frequently criticized as being too time consuming. However, the authors' careful analysis of this part of the study does little more than emphasize the uncontrolled differences between treatments that confound any attempt to interpret the observed time differences. Once again, a fuller explanation of the two treatments would have facilitated unambiguous interpretation of experimental results.

In conclusion, this study provides a valuable contribution to research on the use of CSE in science instruction. The authors were ambitious in characterizing their study as one experiment with three replications. The existence of substantial differences in treatments characterized as identical vitiates this claim. The results of this study certainly suggest, however, that CSE may supplement or replace certain types of laboratory experience in the achievement of particular cognitive and psychomotor objectives. Continuation of this research, with the recommendation that a more rigorous experimental approach be employed, is encouraged.

REFERENCES

Martin, G. R. TIES Research Project Report: The 1972-73 Drill and Practice Study. St. Paul, Minnesota: Minnesota School Districts Data Processing Joint Board, 1973.

Yager, R. E.; H. B. Englen; and B. C. Snider. "Effects of the Laboratory and Demonstration Methods Upon the Outcomes of Instruction in Secondary Biology." Journal of Research in Science Teaching 6: 76-86, 1969.

IN RESPONSE TO THE ANALYSIS OF

Cavin, Claudia S. and J. J. Lagowski. "Effects of Computer Simulated or Laboratory Experiments and Student Aptitude on Achievement and Time in a College General Chemistry Laboratory Course," by Joseph Cotham. Investigations in Science Education, 8(2): 59-65, 1982.

by

Claudia S. Cavin and J. J. Lagowski
The University of Texas

We would like to make the following comments on and clarifications to the abstract and analysis in I.S.E.

Treatments--A description of the individual experiments may be found in Cavin, C. S.; E. D. Cavin; and J. J. Lagowski, "A Study of the Efficacy of Computer-Simulated Laboratory Experiments," Journal of Chemical Education 55: 602, 1978 (reprint available on request) or Cavin, Claudia S., A Study of Some Computer-Simulated Experiments in a College General Chemistry Laboratory Course, Doctoral dissertation, University of Texas at Austin, 1977. The differences in treatment, which may be more meaningful on reading a detailed description, were prompted by the requirements of the two different instructional methods--i.e., laboratory and computer. In the example cited in the abstract, the laboratory experiment involved observation of a color change, the simulation of which would not have been meaningful on a hard-copy computer terminal. In any case, the tests used for comparison related only to those aspects which were common to both methods of performing the experiment.

Student enrollment and scheduling requirements--It was suggested in the analysis that students in specific programs might have scheduling requirements that would influence the laboratory section in which they enrolled. These requirements would seem to be on the basis of time, and it is for this reason that we restricted the study of a particular experiment comparison to a given time period. We had no reason to believe that there was any registration bias for laboratory section in the same time period.

Preece, Peter F. W. "The Concepts of Electromagnetism: A Study of Internal Representation of External Structures." Journal of Research in Science Teaching, 13(6): 517-524, November 1976.

Descriptors--College Science; *Cognitive Processes; *Concept Formation; *Educational Research; *Higher Education; *Physics; Science Education; Scientific Concepts

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Dean Zollman, Kansas State University.

Purpose

Two models of the structure of concepts in electromagnetism were compared with the science graduate students' cognitive structure of the same concepts.

Rationale

The manner in which concepts in science are presented is strongly dependent on the structure of the subject matter. In turn, learning of the concepts depends on the cognitive structure of students. An important question is: Are these two structures similar?

Preece investigated the answer to this question for the basic concepts of electromagnetism. The present work was one of a series of studies on the relation between subject-matter structure and cognitive structure.

As with other similar studies this one was empirical. No model was assumed for the students' cognitive structure or for how the students may have reached the present cognitive state. Instead, data were collected, and results compared to models of the subject matter structure.

Research Design and Procedure

Two models of the relationship among the concepts in electromagnetism were constructed. A spatial model was derived by comparing the dimensions of basic units (length, time, mass and current) for each of the 15 concepts. By assigning a point in a four-dimensional "units space" for each concept, the distances between concepts were established.

The second model, a digraph model, began with electric current as a central concept. Using the defining relations the other concepts were related graphically to the current and/or the other concepts derived from it. A two-dimensional graph of the subject-matter structure resulted.

For each model a matrix of concept proximity was established.

These two models were compared with the cognitive structure of 28 university science graduates who were studying to become physics teachers.

A word-association test asked the subjects to respond with five words most closely associated with each of the 15 concepts of electromagnetism. The data were analyzed by assigning indices of response hierarchy overlap to the subjects' responses. These indices are used as measures of the semantic proximity of the stimuli. A matrix of concept proximity was, thus, established for each student.

For comparision with the models a matrix of mean values was determined from the individual matrices. Correlation coefficients between the mean matrix and each of the model matrices were calculated. Graphical comparisons were also completed for the digraph model.

Findings

A very high correlation (0.80) was obtained between the digraph model matrix and the subjects' proximity matrix. Essentially no correlation (0.24) existed between the spatial model matrix and the proximity matrix.

A graphical analysis of the subjects' proximity matrix yielded a representation of the subject-matter structure which was very similar to the digraph model. This representation contained three distinct groups--current electricity, electrostatics and magnetism.

Interpretations

The digraph model which contained a pivotal concept, electric current, was more closely related to the subjects' cognitive structure than the other model. This result may be an indication that emphasis on central ideas can be useful in teaching science. The results further indicate that a digraph model is a more appropriate method of representing students' cognitive structures.

All but two of the subjects had received degrees in physics. The study supports the idea that students whose cognitive structure matches the experts are more likely to succeed.

ABSTRACTOR'S ANALYSIS

The concepts in any branch of science are related to each other by a structure which develops over many years. While logical development of the concepts contributes most strongly, tradition and history also are involved in determining the subject matter structure. Once this structure is established, textbook authors and teachers seldom deviate from it. In presenting concepts, the structure is followed as if all students already have the necessary cognitive structure and just need to fill in the blanks.

Perhaps, no science is more rigidly structured in its teaching than physics. The vast majority of physics courses and texts follow some general plan in developing concepts. The structure is sufficiently well established that a physics text which deviates from it seldom survives to a second edition. Thus, it is appropriate to compare the structure of physics with students' cognitive structure.

The present paper is one of several recent studies relating science students' cognitive structure to the structure of physics. Previous work, mostly on mechanics concepts, resulted in very similar conclusions. A two-or three-dimensional graphical model of the subject matter structure is able to describe the students' cognitive structure.

The most interesting conclusion of these studies is the form of the structure the students assign to a central location on key concept. All other concepts seem to be, in some sense, derived from this one. Further, the concepts seem to be grouped by the students in a few clusters. Thus, physics teachers could learn about ways to organize and structure teaching.

While the clusters and central concept are interesting, they are at the same time disappointing. In the present study, and in earlier work on mechanics, the central concepts are the ones most emphasized by physicists. The ideas of mechanical force and electrical current are presented many times through physics. Likewise, the clusters represent chapters or sections of a physics book. The material seems to be organized in the students' minds the way it was taught to them. Successful physics students seem to have a cognitive structure very similar to the experts.

The next important question in this research area is: How do the students obtain this cognitive structure? Perhaps, science courses filter out all students whose cognitive structure differs from the "norm." Perhaps, science students enter a science course with a different structure, and structure must be changed to fit the norm. These possibilities have not been addressed sufficiently in any research on subject matter and cognitive structures.

Even more interesting than the cognitive structure of science students is that of nonscience students or beginning science students. Most teachers of science treat these groups as if they have the same cognitive structure as scientists. The courses and textbooks, with a few notable exceptions, are structured the same for all types of students. Only the details such as level of mathematics differs. No rational defense for the assumption of similar cognitive structures can be made.

Some evidence for differences is beginning to emerge. Some recent efforts indicate that students' preconceptions about physics concepts are much different from those of physicists. How these preconceptions fit into the cognitive structure is yet to be determined.

To learn about the cognitive structure of nonscience or beginning science students, research designs different from the present study must be used. Word associations have little meaning for students who do not know the language of science but have heard the same words in other contexts. Comparisons with models acceptable to experts may lead to conclusions that no structure exists when one does.

We must learn how students organize the thousands of everyday experiences related to science. The task will be difficult one for science-oriented researchers to undertake. But, it has tremendous payoffs. We can learn to talk about science to all those people who say they hate it.

IN RESPONSE TO THE ANALYSIS OF

Preece, P.F.W. "The Concepts of Electromagnetism: A Study of the Internal Representation of External Structures" by Dean Zollman. Investigations in Science Education, 8(2): 67-70, 1982.

by

P.F.W. Preece
University of Exeter

I should like to comment briefly on one issue raised by Zollman in his very useful and balanced analysis of my paper (Preece, 1976c). He notes that the cognitive structures of nonscience and of beginning science students are of particular interest, but claims that word-association methods are inappropriate for investigating such students. Although word-association methods are relatively crude instruments for exploring cognitive structures and the development of more sensitive techniques would be very welcome, they have certain advantages particularly for students with little formal experience of the subject matter. Johnson (1969) noted that problem-solving tests were unsuitable for students whose knowledge of concepts could not be expressed as a solution sequence, but that word-association methods did not suffer from this disadvantage. Moreoever, it seems that word associations are particularly sensitive to the early stages of learning (Rothkopf and Thurner, 1970; Shavelson, 1973). Deese (1965) also argued that word-association methods were not contaminated by arithmetical competence or by the rote memorization of examples and were therefore particularly suitable in studies of scientific concepts.

Zollman argues that "word associations have little meaning for students who do not know the language of science but have heard the same words in other contexts." But it is a great merit of the free word-association test that it permits the exploration of the everyday meanings of words, if these predominate in students' memories. Thus the issue raised by Zollman of the possible mismatch between beginning students' preconceptions of physics and the subject-matter structure can be investigated by the word-association technique.

I should like to illustrate this by some research carried out in Exeter (Preece, 1976a). A continued freeword-association test was used to explore the mechanics cognitive structures of nonscience graduates and also of physics students throughout the period of learning that subject at school and university. The cognitive structures of the adult nonscience group and of the 12 year-old beginning physics students were closely similar, these structures reflecting the everyday, nontechnical, meanings of the science concept words. For both groups, three clusters of concepts--kinematics, statics, and energy--emerged in the graphic analyses, although a number of concepts, including <u>distance</u>, <u>time</u>, and <u>work</u>, remained unconnected. For the school and university groups most knowledgeable in physics, all mechanics concepts were interconnected, the three clusters being linked together by the concept <u>force</u>. Further insight into the crystallization of these structures from the concrete, intuitive, and isolated concepts of childhood was provided by multidimensional scaling analyses (Preece, 1976b).

In the Exeter research, the empirical interconcept proximity data were compared with various models of cognitive structure (spatial, hierarchical, and graphic) based on several triads of basic concepts, which defined the dimensions of semantic space or formed the base of a learning hierarchy. For the least knowledgeable groups, the models based on the density-distance-velocity triad fitted the empirical data best, whereas for the most knowledgable groups the mass-distance-time triad gave a better fit. This suggests that density (perhaps through floatation) and velocity (perhaps through overtaking) have a better claim to natural ostensiveness than mass and time, which seem to acquire their special role through the way physics is taught.

Although I have discussed this, and other, research on the organization of scientific concepts in semantic memory in more detail elsewhere (Preece, 1978), I hope that the above brief account is sufficient to show that word-association methods can yield interesting information on the cognitive structures of nonscience and beginning science students.

REFERENCES

Deese, J. The Structure of Associations in Language and Thought. Baltimore: Johns Hopkins Press, 1965.

Johnson, P. E. "On the Communication of Concepts in Science." Journal of Educational Psychology, 60: 32-40, 1969.

Preece, P.F.W. "Associative Structure of Science Concepts," British Journal of Educational Psychology, 46: 174-183, 1976a.

Preece, P.F.W. "Science Concepts in Semantic Space--A Multidimensional Scaling Study." Alberta Journal of Educational Research, 22: 281-288, 1976b.

Preece, P.F.W. "The Concepts of Electromagnetism: A Study of the Internal Representation of External Structures." Journal of Research in Science Teaching, 13: 517-524, 1976c.

Preece, P.F.W. "Exploration of Semantic Space: Review of Research on the Organization of Scientific Concepts in Semantic Memory." Science Education, 63: 547-562, 1978.

Rothkopf, E. Z. and R. D. Thurner. "Effects of Written Instructional Material on the Statistical Structure of Test Essays." Journal of Educational Psychology, 61: 83-89, 1970.

Shavelson, R. J. "Learning from Physics Instruction." Journal of Research in Science Teaching, 10: 101-111, 1973.

Bartov, H. "Can Students Be Taught to Distinguish Between Teleological and Causal Explanations?" Journal of Research in Science Teaching, 15(6): 567-572, 1978.

Descriptors--*Ability; *Biology; *Discrimination Learning; Educational Research; *Instruction; Learning; *Science Education; Secondary Education; Secondary School Students; Teaching Techniques

Expanded abstract and analysis prepared especially for I.S.E. by John Penick, The University of Iowa.

Purpose

This research was concerned with answering two major questions: 1) How well do students distinguish between teleological and causal explanations? and, 2) Can this ability be taught?

Rationale

The author states that the anthropomorphic implications of teleological formulations have been rejected by many science educators. He is concerned that a much more basic aspect of teleology is being ignored by science educators--the danger of confusing ends and the causes by which they are brought about. As Braithwaite (1954) stated, "In a causal explanation the explicandum is explained in terms of a cause which either precedes it or is simultaneous with it; in a teleological explanation the explicandum is explained as being causally related either to a particular goal in the future or to a biological end is as much future as present or past."

The author's basic assumption is "that in order to achieve the ability to distinguish between causal and teleological explanations, the student had to learn one simple principle, namely that biological processes are not brought about by their ends but by specific causal, mainly neural and hormonal, mechanisms." Bartov continues his premise by stating "that the quite simple principle in question, which can be exemplified by a great number of otherwise unrelated biological facts, may be readily learned, well retained, and subsequently applied by most high school students." With this in mind, an experiment was set up to assess students' ability and to measure their gain.

Research and Design Procedures

Six hundred tenth-grade science students were given pretests, and 390 completed both the pretests and the posttest in a nonequivalent control group design. These students were from 21 classes in five high schools; 13 in the experimental group and 8 classes in the control group.

All of these classes studied biology with their normal classroom teacher with the experimental classes receiving an additional five lessons. The control classes received no special treatment and the regular biology lessons were claimed to have no effect on the investigated question.

The treatment lessons were also conducted by regular classroom teachers who followed study guides. Three of these lessons dealt with biological principles and two of the lessons were devoted to discussion of homework assigned at the end of the other treatment lessons. These treatment lessons involved a particular biological phenomenon about which students would suggest hypotheses, draw deductions from the hypotheses and plan and perform experiments to verify deductions.

The test instrument, Test On Causal And Teleological Relationships (TOCATR), provided students with a statement of a biological phenomenon and four sample statements: one each of causal, teleological, anthropomorphic, and both teleological and anthropomorphic explanations of the phenomenon. Scoring was accomplished by giving one point for each correct response.

Pretest scores were used as a covariate in analysis of covariance with posttest scores being criterion variables. In addition, another version of the TOCATR Test was used in a posttest only form. This version's data were analyzed using a t-test between experimental and control groups.

Findings

Pretest scores supported the assumption that secondary school students have difficulty discriminating between causes and purposes. Analysis of covariance indicates that at the end of the treatment the experimental group had significantly higher scores than the control group. T-test results of a comparison of posttest only TOCATR No. 2 scores indicated the same significant difference in favor of the experimental group.

Interpretations

The author concludes that the experimental treatment of five lessons was effective in improving the ability of secondary school students to distinguish between teleological and causal explanations. He sees this as justification for providing special treatment to students developing these abilities and recommends that lessons be conducted according to specific suggestions in the paper and that in any discussion, students should be asked to provide both causal and teleological explanations. As a final note, Bartov suggests that questions similar to the TOCATR should be included in examination papers in biology for a routine evaluation of student achievement.

ABSTRACTOR'S ANALYSIS

Bartov has attempted to investigate a relatively complex area. Just as generations of biologists and philosophers have misunderstood Aristotle's use of teleology, leading to violent rejection of Aristotle by Bacon, Descartes, and others, modern day biology students are equally unlikely to be able to differentiate between causes and purposes.

Bartov's development of an instrument to assess this difficulty and his design of a simple and short experimental program to alleviate the deficiency are commendable. It is self-evident that teachers should be concerned with these differences, but, I suspect, few teachers deal with the philosophy of science, much less worry about how students phrase statements of cause and purpose during explanations.

Since philosophers of science have had great difficulty in dealing with teleological principles, I must take exception with Bartov's statement that "the quite simple principle in question . . . may be readily learned, well-retained, and subsequently applied by most high school students." Certainly, if this were the case, then his pretest scores would have been considerably higher and the study itself would be unnecessary.

The study is weakened somewhat by the nonrandomness and inequality of the experimental and control groups. Adding to this difficulty is the fact the experimental groups got the same lessons as the control group plus an additional five special lessons designed to specifically deal with suggesting hypotheses, deducing from hypotheses, planning and performing experiments, and drawing conclusions. Certainly you would expect the group who got everything plus something else to have gained from the experience. The author states that the regular lessons were irrelevant to the questions being investigated but provides no evidence of that. Even if the lessons themselves did not contain relevant information, no observations of teacher performance were made and, thus, we have absolutely no notion of what individual teachers did in the privacy of their own classrooms.

The treatment itself sounds most interesting and innovative. It would be very useful to the reader if the treatment were explained somewhat better so that the reader could know precisely what happened in the classroom. It is unclear what went on during the lessons which were conducted after a sequence where apparently students suggested hypotheses and worked with them. The study could not be replicated from the information given, by any means.

The Test On Causal And Teleological Relationships instrument sounds most interesting. I could certainly agree with the author that such items would be appropriate for partial evaluation of student achievement in biology. As interesting as the test instrument is, however, it is certainly not appropriate to call it a "standardized test" as does the author.

Analysis of covariance is quite reasonable to use in this case and showed that the covariate (the pretest) showed a significant difference between the two groups. With this in mind, it would now seem inappropriate to do an analysis of the posttest only scores. The author also erroneously states that "only a t-test could be and was used to compare the mean scores of experimental and control groups." Certainly, there would seem to be a number of other appropriate statistical tests which could be used.

The statement, "the experimental group made substantially more progress than the control" cannot be legitimately made since the comparison that he ran with this analysis of covariance was a posttest comparison and not a pretest/posttest comparison.

The written report could be improved greatly by expanding the explanation of the treatment, more careful wording of the paper in general, and by providing more specific implications about what to do with the information we have learned from this study. For instance, Bartov states that "lessons on reflexes and tropisms should be conducted according to the suggestions outlined in this paper" is not a very useful statement since the suggestions were not really outlined.

Very little research is done on the effect of teaching the philosophy of science in high schools and it would seem to be a very fertile area for the future. It would be interesting to find out if Bartov's instrument detects significant differences among different principles within science, differences between teaching strategies, or differences between various curricula when teaching is controlled.

REFERENCE

Braithwaite, R. D. Scientific Explanation. Cambridge: Cambridge University Press, 1954.

IN RESPONSE TO THE ANALYSIS OF

Bartov, J., "Can Students be Taught to Distinguish Between Teleological and Causal Explanations?" by John Penick. Investigations in Science Education, 8(2): 73-78, 1982.

by

H. Bartov
Hebrew University

The quotation from Braithwaite (1953), defining the notions of teleological and causal explanations, is very confusing and must be dealt with in some detail.

The statement "In a teleological explanation the explicandum is explained as being causally related either to a particular goal in the future or to a biological end which is as much future as present or past" seems to obliterate or even to annul the differences between teleological and causal explanations.

Braithwaite's distinction between "goals" and "biological ends" is highly individualistic and of little or no importance in biology, at any rate--in high school biology. According to Braithwaite "the peculiarity of a biological end is that it is a permanent goal," like the continuous beating of the heart responsible for the continuous circulation of the blood, and therefore biological ends are "as much future as present or past." Whereas goals refer to a particular short action, such as a single beating of the heart, "responsible for the circulation of blood a short time afterwards."

When only goals are taken into consideration, as it should be, the differences between teleological and causal explanations are quite distinct and clear. "A teleological answer (to the question 'Why') explains a present event by means of a future event, ...a nonteleological answer--in terms of a present or past cause." This rule holds good for "all teleological explanations which are not reducible to explanations in terms of a conscious intention to attain the goal" (Braithwaite, 1953).

Moreover, even intentional activities (which are limited to human beings only--at least in the opinion of most biologists) are no exception to the above rule. "Teleological explanations of intentional goal-directed activities are always understood as reducible to causal explanations with intentions as causes (Braithwaite, 1953). In the case of intentional activities the action is not caused by the yet unattained and--maybe never to be attained--future goal, but by the desire to attain it, which precedes the action. Intentional explanations conform to the conditions for causal explanation, and should be regarded as such explanations, to all intents and purposes. Teleological (nonintentional) explanations refer to future goals and not to present or past causes.

The definitions of teleological and causal explanations were discussed at some length, because of their crucial importance for the issue under consideration. Some other points raised in the review are dealt with briefly in the following lines.

The reviewer questions the hypothesis, that "the quite simple principle in question (that biological processes are not brought about by their ends) may be readily learned, well retained and subsequently applied by most high school students."

This question may be answered only by experiment, such as that described in the reviewed article.

I certainly admit that "the study could not be replicated from the information given (in the article), by any means." The reviewer's suggestions for further research, such as "to find out if Bartov's instrument detects significant differences among different principles within science, differences between teaching strategies," etc., were also made by me in my doctoral dissertation (written in Hebrew). I specifically referred to the so much debated question of the relative efficacy of the discussion, demonstration, and laboratory methods in teaching biology. The treatment lessons developed for my study can be given by any of these methods. They have also the advantage of economy of time, as only five lessons are needed to accomplish the treatment. I would be very happy to have the possibility to collaborate in improving and replicating my study under discussion in more controlled conditions than were possible here in Israel, as well as in planning and carrying out new investigations, based on the tests and treatment lessons developed for the study described in the reviewed article.

REFERENCES

Braithwaite, R. D. Scientific Explanation. Cambridge: Cambridge University Press, 1953.