DOCUMENT RESUME

ED 179 425 SE 029 579

AUTHOR	Blosser, Patricia E., Ed.; Steiner, Robert L., Ed.
TITLE	Investigations in Science Education, Vol. 6, No. 1. Winter 1980. Expanded Abstracts and Critical Analyses of Recent Research.
INSTITUTION	Ohio State Univ., Columbus. Center for Science and Mathematics Education.
PUB DATE	80
NOTE	81p.
AVAILABLE FROM	Information Reference Center (ERIC/IRC), The Ohio State University, 1200 Chambers Rd., 3rd Floor, Columbus, OH 43212 (subscription $6.00, $1.75 ea.)
JOURNAL CIT.	Investigations in Science Education; v6 n1 Win 1980
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Abstracts; *Cognitive Development; *Educational Research; *Instruction; *Learning; Research; Research Methodology; Science Education; *Scientific Literacy

ABSTRACT

Presented are analytical abstracts, prepared by science educators, of nine articles which report research in the areas of cognitive development, learning, instruction, and scientific literacy. Each abstract contains bibliographical data, research design and procedure, purpose, research rationale, and the abstractor's critical analysis of the research report. (PB)

ED179425

INVESTIGATIONS IN SCIENCE EDUCATION

National Association for Research in Science Teaching

ERIC Clearinghouse for Science, Mathematics,
and Environmental Education

Published Quarterly by

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

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

SE 029 579

Volume 6, No. 1, 1980

NOTES FROM THE EDITORS

This issue of INVESTIGATIONS IN SCIENCE EDUCATION contains the critical abstracts and analyses of nine articles. Although these nine articles have not been clustered, there are some common variables. Several relate to cognitive development and/or learning: Allen, Bredderman, Lawson, and Scott. Others focus on instruction: Anderson and Lee, Hibbard and Novak, Parker and Mertens, and Venderbroucke. One relates to scientific literacy: Doran.

Also included are three responses to articles previously published in I.S.E. We are pleased that the authors have provided additional information about their research and that they have attempted to answer or respond to comments raised in the abstractor's analysis section of the paper reviewing their published research. We hope that other authors will feel free to respond to or comment about their article as reviewed for I.S.E.

Patricia E. Blosser
Editor

Robert L. Steiner
Associate Editor

Allen, Leslie R. "An Examination of the Ability of Third Grade Children from the Science Curriculum Improvement Study (SCIS) to Identify Experimental Variables and to Recognize Change." Science Education, 57(2):135-151, 1973.

Descriptors--*Cognitive Development; Curriculum; Educational Research; *Elementary School Science; *Instruction; *Longitudinal Studies; Science Course Improvement Project; *Science Education

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

Purpose

The purpose of the study was to compare the performance on specific tasks related to some SCIS objectives of third graders who had participated in a SCIS Program versus third graders who had not. Specifically, the children were compared on their ability to:

1. identify variables in a system
2. identify changes in a system

The research hypotheses tested in the study but not explicitly stated are as follows:

1. There will be no differences between the groups in their ability to identify variables in a system.
2. There will be no differences between groups in their ability to identify changes in a system.
3. There will be no differences between male and female pupils in their ability to identify variables in a system.
4. There will be no differences between males and females in their ability to identify changes in a system.
5. There will be no differences between socio-economic status and the ability to identify variables in a system.
6. There will be no differences between socio-economic status and the ability to identify changes in a system.

Rationale

This study is but one of a series written by the author and based on a longitudinal study conducted to determine the effectiveness of the SCIS program on children's acquisition of specific scientific concepts, process skills, and positive attitudes toward science. The author is attempting to assess the effect of the program on children's intellectual development. As the children proceed through the program, they are evaluated on the basis of the program's stated objectives. At this point in time four studies have preceded this report. Thus, the study was an extension of previous studies and it appears that this research will continue for the next three years (grades 1-6).

Research Design and Procedure

When the initial study was begun some three years ago, 50 children were randomly chosen from three socio-economic levels (upper, average, and lower). Thus, a total of 150 pupils participating in the SCIS program from three socio-economic levels was chosen. Next, schools having similar children but not participating in the SCIS program were selected and again 50 pupils from each of the three socio-economic levels were chosen. This made a grand total of 300 pupils who participated in the longitudinal study conducted up to this point in time. However, because of mortality, only 176 pupils were available for this particular study reported. The male-female ratio was just about 50/50 with an equivalent number of children representing each of the three socio-economic levels. All children were from the third grade, with half participating in the SCIS program continuously over a period of three years. The children were evaluated at the conclusion of the third academic year.

The experimental design used is the posttest-only control group design as defined in Campbell and Stanley's article. At the completion of the third year, after having studied the SCIS Physical Science Unit, all children were tested individually "in a room set aside for this purpose or at a lanai table." All children were evaluated on five test items presented one-at-a-time. Four of the items required the children to identify changes that took place when a candle was lit with a wooden match and covered by a glass beaker. The items were designed so that children could manipulate objects and explain their actions. Responses were recorded for the children on each of the items presented.

Children who were not participants in the SCIS program were exposed to some science instruction. The author reports that their science program consisted of a science text entitled Concepts in Science which placed heavy reliance on verbal interactions rather than the physical manipulation of objects. In other words, it was not a science inquiry program, but one in which children were taught elementary science using the traditional methods so prominent during the first 60 years of this century.

The author summarized the number of pupils and their sex in the three socio-economic levels used in the study. Thus, one can see at a glance the total number of pupils: both SCIS and non-SCIS, their sex, and their division within the three socio-economic levels identified.

As for the findings of the testing undertaken at the close of the school year (May-June, 1977), the author notes that in a previous study the five test items used loaded heavily on a "cognitive factor" using factor analysis methods. The author goes on to say that, "in the interest of clarity," a separate analysis of each item will be presented using ANCOVA methods. The California Test of Mental Maturity (CTMM) was used as the covariate. It is also reported that the average correlation for each item and its scale total based on factor loadings was 0.71.

There are several tables in the article which summarize the testing results. Tasks on which the data are summarized are as follows:

1. Variables effecting time taken to transfer water using an eye dropper.
2. Variables effecting level at which a plastic syringe floats.

3. Variables effecting the distance a stone is thrown by a slingshot.

4. Variables effecting the distance travelled by a toy truck.

5. Changes that occur during an experiment. This was further subdivided into three parts:

   a. changes observed in a burning candle
   b. changes observed in the match and match-box
   c. changes observed when a beaker is lowered over a burning candle

Analysis of covariance for the factors of science program, sex, and socio-economic level are presented for the five tasks. Mean scores, standard deviations, and t-tests for the three socio-economic levels on four of the five tasks are reported. In addition, summaries of the children's responses to the five tasks are listed and categorized into SCIS - non-SCIS and socio-economic levels.

Findings

The ANCOVA revealed that there were significant differences between SCIS children and non-SCIS children on all five of the tasks. The difference was in favor of the SCIS children. Three of the analyses revealed that there was a difference between males and females in favor of the boys on tasks 3, 4, and 5. With respect to socio-economic level, analyses revealed that there was a significant difference. T-tests indicated that, on three tasks, the upper socio-economic group outperformed the other two socio-economic groups while the average group outperformed the lower group on one of the tasks. No significant interaction effects were found or reported.

Interpretations

The author concludes that, on the basis of the findings reported in the study, Honolulu third grade pupils who had participated in the

SCIS program are significantly better than third grade pupils enrolled in a traditional science program on their ability to identify variables and changes in systems.

ABSTRACTOR'S ANALYSIS

Several questions are raised after reading the study. The introduction is very skimpy. While it does refer to a continuous evaluation of the SCIS program, it does not place the present study in proper perspective relative to the other studies referred to in the bibliography. What is the overall long-range objective of the longitudinal study? How does this study relate to them and the overall longitudinal study?

The purpose of the study is straight forward. However, no mention is made as to why evaluation of the SCIS Physical Science unit is made and not the Living Science unit. Why not evaluate the program after the pupils have experienced both units? What significance does the Physical Science unit have in relation to the Living Science unit? The author notes, too, that "certain SCIS objectives are considered." This implies that there were others and that these were chosen instead. Why were these objectives covered and not others? One can be led to a conclusion that maybe additional publications can be produced by using this procedure of reporting. Clearly a problem statement should be included and a rationale as to its importance or significance relative to science education research.

A main criticism of the study lies in the research design. No research hypotheses are ever stated. Isn't the purpose of statistical analysis designed to test hypotheses? These should be stated so that the reader does not have to infer what they are.

While the experimental population for the study is identified, no real information is provided here in the study. The reader is referred to previous studies for this information. Really good research does not use this type of reporting procedure. The reader should not have to review other studies in order to understand the present study. The population should be identified and described in detail. In addition, the author does not identify the independent and dependent variables. However, they can be inferred from reading the study.

Further, no description of sampling procedure for choosing schools and children is described. Why did certain schools have the SCIS program and others, not? This is not made clear to the reader. What controls were made or introduced to control for teacher effects? In a study such as this, one would want to be able to control for this factor in order that adequate comparisons could be made. This has to be done so that one can be sure that differences found can be attributed to the treatment only or else the results may be called into question.

Concerning the treatment, little is said about this crucial factor. There is no mention of how long the control group has been exposed to science instruction of any sort. In addition, time and duration of science for the third graders is not reported. Also, how many teachers were involved in the instruction of science for both the experimental and control groups? What is their background and other pertinent information related to instructional methods? These are important questions that should be answered in order to evaluate the effectiveness of the SCIS program.

Only one sentence is devoted to a description of the testing procedures used. The tasks used to make the assessment are adequately described. But, who performed the testing? How long did it take to test the children? Was there any method used to test the accuracy and reliability of the responses recorded? The reliability and validity of the tasks are not adequately described here. It is reported that they were piloted with grade three children. But one is left hanging in terms of a description for this process. This needs to be explained further and additional information provided here for the reader.

The ANCOVA tables summarize the data very neatly. As was pointed out by the author, the scores could have been combined and one analysis summarized for all SCIS and non-SCIS childrens' performance on the tasks. It was a good move on the part of the author to provide analyses and summaries for all five tasks. This procedure does aid the reader in understanding some of the things that are being considered in the evaluation even though it results in reading some 17 tables in all. The author does summarize the data for means, standard deviations, and t-tests by socio-economic level in tables for four of the five tasks presented. Why doesn't he do the same for the fifth task? No explanation is offered for this omission. In addition, the author states that two of the analyses show scores in favor of boys, yet a critical review of the table reveals three such significant differences in favor of boys. An alternate explanation here may be that the girls outperformed the boys on one of the tasks. But this remains unanswered in the analyses.

It is also reported that the upper socio-economic group outperformed both the average and lower socio-economic groups on three of the tasks and the average group outperformed the lower socio-economic group on one of the tasks. But no summarization is presented for the fifth task. The author makes no note of this. Perhaps there were no significant differences. But that could have been stated in the findings without including a table.

After the summary of the findings the author concludes that Honolulu children in the SCIS program are superior to children not in a SCIS program on ability to identify variables and changes. But no appropriate interpretations or generalizations are made. Is this because the focus of the study is so narrow? Certainly the implications of the study and its limitations should be discussed. Where does it fit into the matrix of similar kinds of elementary science program evaluations? Does it make a contribution? Does it shed some light on any of the problems that are the major concern of science education research? Or, another way of posing this question is: Does it reduce, clarify, or throw any new light on the original area or problem of concern? Does additional research need to be conducted, and what direction should it take? Unless some of these questions are answered, it is difficult to make any judgment about the results of the study. Certainly in reporting research of this nature, questions such as those listed above should not be left unanswered.

Anderson, O. Roger and Mae T. Lee. "Structure in Science Communications and Student Recall of Knowledge." Science Education, 59(1):127-138, 1975.

Descriptors--Educational Research; *Instruction; Learning Theories; *Retention; *Science Education; *Secondary School Science; Stimulus Behavior; *Verbal Communication

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Willis J. Horak, University of Arizona.

Purpose

The purpose of this study was to investigate the relative relationships between amount of science content structure in lessons and the amount of knowledge retained after instructional communication. Additionally the study was designed to allow analysis of fine variations in knowledge acquisition. These fine variations were then related to the overall structure of the statements or discourse units containing the communicated content.

Rationale

The bio-psychological theory of structure in human communication developed by O. Roger Anderson and presented in his books Structure in Teaching: Theory and Analysis and Quantitative Analysis of Structure in Teaching formed the theoretical framework for this study. Basically, this theory holds that the effectiveness of human communication is biased by the facts that (1) the environment favored the evolution of organisms possessing receptors sensitive to periodic or repetitive stimuli, (2) the presence of periodic stimuli in space and time have induced, through natural selection, the appearance of highly advanced forms of living organisms capable of exploiting the environment by utilizing information in periodic stimuli, (3) the dependency on periodic stimulation, such as light energy and movement of surrounding matter, has induced a perceptual bias to readily assimilate repetive stimuli, thereby producing changes in behavior, and (4) the perceptual bias favored by organic evolution and reinforced by periodic visual stimuli during ontogeny produces communication patterns in higher animals containing repetitive sounds (Anderson, 1970).

This theory may be interpreted in light of the results of research on attention and arousal which maintains that too much repetive patterned input can produce lack of attention. Thus, lessons need some novel stimulus onset or repetitive removal to cause continued readiness to receive and to encode sensory input. Optimum communications would consequently "contain a pattern of integrated discourse introduced by a novel stimulus which serves as an arousal step to prepare the individual to receive some subsequent communicated content."

Research Design and Procedure

This study was conducted in two phases. Subjects for the first phase consisted of 61 female ninth and tenth grade students in a private urban secondary school. Subjects for the second phase, which was viewed as a test of generality of the findings, consisted of 41 male and female students in an urban junior high school. Separate treatment communications were constructed for each phase of the investigation. The topics chosen for the studies dealt with "African Sleeping Sickness" and "Life in the Ocean" respectively. These topics were reasonably unfamiliar to the students and hence prior knowledge should not have interfered with the analysis.

For each phase two treatment conditions were developed. The treatments consisted of alternating high and low structure spans as ascertained by kinetogram analysis of the discourse units. One treatment condition consisted of a sequence, identified as "high-low," containing alternating spans with the odd spans high structure and the even spans low structure. The other treatment condition consisted of a sequence, identified as "low-high," containing alternating spans with the odd spans low structure and the even spans high structure. The amount and kind of information within each span remained constant thus allowing direct comparison of the high and low structure spans.

Audiotape recorded messages were the instructional mode of communication and the criterion measure consisted of a free recall response. Immediately after instruction students were asked to list in writing as many of the ideas or statements as it was possible for them to remember. The frequency of recall of the items was plotted on a histogram and compared to the kinetogram graph of the communicated lesson structure. Total scores were utilized to compare the mean high-structure recall scores with the mean low-structure recall scores. Statistical significance of the differences in the means was analyzed utilizing a two-tailed t-test.

Findings

The amount of structure in a communication was found to be directly related to the amount of information recalled after hearing a communication. When the kinetogram sloped upward or remained in an elevated position denoting increased structure, there was a corresponding increase in the number of items recalled as evidenced by the histogram. In five of the six spans analyzed high structure spans produced statistically significant ($p < .05$) more recall than low structure spans.

Interpretations

The findings support the theory of human communication outlined previously by O. Roger Anderson. The fact that the theory predicted the observed outcomes strengthens and enhances its validity. Additionally, the two phases of this study showed that science content and student grade level—junior or senior high school level—did not affect the overall superiority of high structure lessons. One point that needs further study, however, is the length of spans and their relationship to knowledge acquisition. This is evident from the fact that in this study one of the longer spans of discourse utilized failed to show significant differences between high structure and low structure treatments. The results of this study may also be interpreted to indicate that science teachers should "give consideration to the amount of structure in their communications particularly when the students are expected to recall and apply this information in situations that are openended or provide minimum contextual cues to aid recall."

ABSTRACTOR'S ANALYSIS

One cannot reflect on teaching without realizing that teaching involves communication. Consequently it is evident that any educational theory relating communication and learning should be most useful for educators. Communication and structure in teaching has been researched from many viewpoints. Gagne's learning hierarchies, Ausubel's advance organizers, Schwab's philosophical analyses of content knowledge, Gutman's psychological viewpoints of structure and Bellack's work with teaching episodes or teaching cycles are just a few of the diverse studies that come to mind when one considers structure in teaching. However, none of these research areas address themselves as pointedly to indepth analysis of teacher communicated subject matter structure and the quantitative determination of lesson structure as does O. Roger Anderson's kinetic structure analysis. Thus, research such as the report reviewed here should reveal new insights into the processes of symbolizing that constitutes a large part of everyone's schooling.

Although the theoretical framework for studies of this type was developed and published approximately nine years ago, the area has not yet been thoroughly researched. However, there are indications that the theory is being expanded and integrated into some long-range research programs. One of the weaknesses of the reported studies appears to be the narrow view of achievement as knowledge acquisition. In many of the early studies the criterion measures were based on factual item or discourse unit or substantive work identification. This present (reviewed) study also measured knowledge level responses by assessing free recall of specific communicated facts. More recent studies such as those reported in the Journal of Research in Science Teaching by Mathis and Shrum (1977) and Ferraro, Lee, and Anderson

(1977) have expanded the theory to include broader achievement definitions and more diverse sample or treatment groups.

Some research evidence exists which supports the contention that other types of teaching structure do differentially affect knowledge and application level achievement. Peterson (1977) utilized ATI methods and found differential achievement effects when the criterion measures were essay and multiple choice tests, the treatments involved structure, and the aptitude measured was manifest anxiety. Similarly, Horak and Slobodzian (1978) utilizing ATI methods, lesson structure variables, and an aptitude measure of locus of control, found differential achievement effects when the criterion measures concerned science content or science processes. These other types of school achievement may be differentially affected by the kinetic structure of the communicated discourse. Thus results of the kinetic structure research need to be expanded into areas such as these if it is to prove most valuable for all educators.

In a recent educational article, Glaser (1972) pointed out that instructional treatments must be generated by a systematic analysis of the kinds of psychological processes called upon in each particular instructional method and achievement measure. This admonition may be most appropriate for kinetic structure research. The fact that external organization of the science information is lacking in low kinetic structure communications may force students to utilize other psychological processes in encoding the relevant information. Retention studies may provide some insights into this area.

This study adds significantly to the rationale underlying the basic kinetic structure theory. It not only expands the generalizability of the theory to other science content areas, but also to other grade levels. These results have been further extended by the more recent articles previously cited (Mathis and Shrum, 1977; Ferraro, Lee, and Anderson, 1977). Additionally this research study has provided a method for analyzing fine variations in knowledge acquisition by developing communications that alternated high and low structure spans. The range of the mean structure coefficients, however, appears to vary substantially between the two treatment communications. For the low-high treatment the reported mean fundamental coefficients vary from a low of 0.23 to a high of 0.48. For the high-low treatment the reported mean fundamental coefficients vary from a low of 0.13 to a high of 0.52. This differential treatment structure may have affected the results of the t-test analysis in an undetermined way.

In this research report it is stated that, in a sense, each subject served as his or her own control for extraneous variables since the performance of the subjects with high structure spans was compared against their performance with low structure spans. From the report, it is questionable whether this is exactly true since, for example, subjects with a high structure treatment on span two were compared with different subjects with a low structure treatment on span two. This was true for all six of the analyzed spans. The number of discourse units in each span was not equalized to allow such a direct comparison.

The description of the underlying theoretical base is most helpful. The extension of the basic theory based on neurophysiological concepts is also very useful. This discussion is quite in-depth and therefore quite informative for readers not entirely familiar with the presented theory. More studies need to present this type of in-depth rationale. The identification and explanation of the structure coefficients was concise and understandable. However, a more thorough explanation of the scoring procedures of the free recall test would have been helpful. One wonders if there is a relationship between the number 10.6 reported for low structure span three and the number 7.1 reported for high structure span one. If a student recalled a word that is a substantive word in four discourse units in a span does this word get tabulated in one or in four discourse units? One example of a scored span would have helped immensely. Additionally, this may have been clarified by reporting the number of discourse units in each span.

Overall this research report is very thorough. It is well written and understandable. People not familiar with kinetic structure analysis of communications can profit from the overview of the theory that is presented. A further analysis of reports cited and instructional methods employed in the design of the experimental research may reveal that basically kinetic analysis of communications may have a large input on individualized programmed learning. It may also affect future textbook format and overall curriculum design. More teachers need to be aware of these findings related to knowledge acquisition and communication structure.

REFERENCES

Anderson, O. R. Structure in Teaching: Theory and Analysis. New York: Teachers College Press, 1969.

Anderson, O. R. "A Comparative Analysis of Structure in Teacher Communicated Science Content." Journal of Research in Science Teaching, 7:227-244, 1970.

Anderson, O. R. Quantitative Analysis of Structure in Teaching. New York: Teachers College Press, 1971.

Ferraro, E.; M. T. Lee; and O. R. Anderson. "The Effects of Structure in Science Communications on Knowledge Acquisition and Conceptual Organization by Students of Varying Mental Maturity." Journal of Research in Science Teaching, 14:441-447, 1977.

Glaser, R. "Individuals and Learning: The New Aptitudes." Educational Researcher, 1:5-13, 1972.

Horak, W. J. and K. A. Slobodzian. "The Effects of Locus of Control and Method of Presentation on Pre-Service Teacher's Achievement." Paper presented at the 51st annual meeting of the National Association for Research in Science Teaching, Toronto, Canada, 1978.

Mathis, P. M. and J. W. Shrum. "The Effect of Kinetic Structure on Achievement and Total Attendance Time in Audio-Tutorial Biology." Journal of Research in Science Teaching, 14:105-115, 1977.

Peterson, P. L. "Interactive Effects of Student Anxiety, Achievement Orientation and Teacher Behavior on Student Achievement and Attitude." Journal of Educational Psychology, 69:779-792, 1977.

Bredderman, Theodore A. "The Effects of Training on the Development of the Ability to Control Variables." Journal of Research in Science Teaching, 10(3):189-200, 1973.

Descriptors--*Cognitive Ability; Doctoral Theses; *Elementary School Students; *Intellectual Development; *Learning Theories; Psychology; *Retention Studies; Science Education; Training

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Donald E. Riechard, Emory University

Purpose

The major purposes of this study were:

1. To determine if the ability to control variables could be acquired at an earlier age by students receiving relatively extensive formal training.
2. To assess the relative effectiveness of external reinforcement and cognitive conflict-based treatments.

Rationale

The study is grounded primarily on Piagetian research. Inhelder and Piaget (1958) reported that children typically were unable to control variables until 14 or 15 years of age. The present study sought to determine if that age level could be lowered.

Piaget's equilibration theory provides rationale for the second purpose defined above. Furth (1969) describes the theory in writing, "The essential point in the 'knowing circle' is the internal structure. The circle assimilates or incorporates the real event into the structure and at the same time accommodates the structure to the particular features of the real event." The emphasis is on internal conflict which upsets existing cognitive equilibrium. Learning results when equilibrium is restored through accommodation of internal cognitive structures. The equilibration model is contrasted with other theories which place a high value on external reinforcement.

Research Design and Procedure

The subjects of this study were 27 fifth and sixth grade students from a single school. The population consisted of all of the students in the two grades (230) less a special class of slow learners. Students were randomly selected from the population and pretested until 27 who could not control variables were identified. The 27 subjects were partitioned evenly (9 each) among three groups: control, external reinforcement, and internal cognitive conflict.

There were three test administrations: the pretest (mentioned above), the posttest, and a retention test. Based on information given in this report, the design can be diagrammed according to Campbell and Stanley (1963) as follows:

$$
\begin{array}{cccc}
0 & X_1 & 0 & 0 \\
0 & X_2 & 0 & 0 \\
0 & & 0 & 0
\end{array}
$$

A different test form was used at each administration (observation). The forms were all patterned after those used by Inhelder and Piaget (1958) and were identified as "rods," "springs," and "levers." Each provided a situation in which five independent variables and one dependent variable were operating.

During the treatment phase, students in the reinforcement and conflict groups received training individually. There were four treatment sessions for each group.

The <u>reinforcement</u> treatment focused on having the student repeatedly carry out 31 properly controlled experiments. The predominant tactic was to talk the student through a series of controlled experiments so as to reveal a consistent pattern of variation between the dependent and one independent variable. The student was then asked to predict the effect on the dependent variable of a particular change in the one independent variable.

For the treatment designed to induce conflict the student was led through a sequence of 31 experiments in which two or three independent variables were changed in each experiment. However, the change in one of these variables and the dependent variable followed a consistent pattern giving misleading support for establishing an invalid relationship between them and for ignoring other changed variables.

Findings

At the time of the retention test, nearly half of the subjects were controlling variables. The mean age of the subjects was 11.8 years.

The post and retention test means for the three combined groups were found to be significantly different ($p < .01$) from the pretest means. However, while the mean scores were slightly greater for groups receiving training, the differences among the groups on any of the three tests were not significant (alpha = .05).

Interpretations

The author concluded the following:

1. Numerous experiences with problems requiring the controlling of variables can accelerate the development of this ability for some fifth and sixth grade children.

2. Conversely, some children's development of the ability at this age may be unaffected by formal training.

3. It makes little difference whether the training involves experiences including external reinforcement, the inducement of cognitive conflict, or merely the repeated posing of the problem to be solved.

ABSTRACTOR'S ANALYSIS

Validity of the Study. The author seems quite safe in concluding that external reinforcement might not be necessary for behavioral change to occur. And such a finding is supported by the work of Smedslund (1961).

However, the fact that <u>all</u> groups (cognitive-conflict, external reinforcement, and control) made essentially the same gains raises some question about what <u>is</u> necessary for behavioral change. Could the gains have been the result of what Campbell and Stanley (1963) call maturation, history, testing, or interaction of testing and X? The author insightfully suggests a problem with testing. A Solomon Four-Group design would have been most appropriate here.

The question on maturation/history is still of interest, however, since the length of the study is not given in the report. Not only were the posttest means for all groups significantly higher than pretest means but all groups performed slightly better on the retention test than on the posttests. How long were the periods between pretests, posttests, and retention tests?

Other questions might be raised about the three test forms ("rods," "springs," and "levers") used in determining the subjects' abilities to control variables. Of major concern is whether or not the three are equivalent forms. Further, the order in which the forms were administered is not entirely clear. One can only infer that "rods" was the pretest, "springs" the posttest, and "levers" the retention test.

The author found that at the retention test nearly half of the subjects were operating at the formal operational level (substage IIIB). The mean age of this group was 11.8 years. While that age is slightly lower than found by Piaget and his associates, it is of interest that other researchers have found formal operations to develop several years <u>later</u> than 11.8 (Renner and Lawson, 1973).

Research Design. The research design reported for this study is adequate. However, the description of the sampling technique leaves some questions unanswered. It is stated that the sample was drawn from 230 fifth and sixth grade students. However, "a special class of slow learners was excluded from the population to be sampled." The number of students in the slow learners class was not given. Thus, what was the size of actual population?

The exact method of sampling is also unclear. Were 31 students randomly selected and then that group pretested "until 27 who did not control variables were identified?" More likely, students were randomly selected one at a time from the population and pretested until the 27 students who would participate in the study were identified.

Finally, the author does not indicate that the 27 students were randomly assigned to the three groups. If random assignments had been made (or reported if they were made) the research design diagrammed above would have been strengthened greatly.

Suggestions for Written Reports. This written report conveys its message quite well. The abstractor feels, however, that the report suffers from a very common problem. Namely, it is the problem of condensing a relatively large study (in this case a doctoral thesis) into a journal article. In this article, for example, the introduction and discussion seem to include much more than is needed to adequately inform the reader. Excess material can often make an article difficult to follow.

What to include and what not to include in a journal article become critical decisions. And these decisions are especially difficult for the author of the larger piece (e.g., the thesis) to make; he/she is often too close to the study. It would, therefore, be wise for him/her to have a colleague not connected with the study to review the journal piece for clarity and preciseness.

REFERENCES

Campbell, D. T. and J. C. Stanley. "Experimental and Quasi-Experimental Designs for Research on Teaching," in N. L. Gage (ed.), Handbook of Research on Teaching. Chicago: Rand McNally and Co., 1963, pp. 171-246.

Furth, H. Piaget and Knowledge. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1969.

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

Renner, J. W. and A. Lawson. "Piagetian Theory and Instruction in Physics." The Physics Teacher, 11:165-169, 1973.

Smedslund, J. "The Acquisition of Conservation of Substance and Weight in Children. V. Practice in Conflict Situation Without External Reinforcement." Scandinavian Journal of Psychology, 2:156-160, 1961.

Doran, Rodney; Robert O. Guerin; and Joanne Cavalieri. "An Analysis of Several Instruments Measuring 'Nature of Science' Objectives." Science Education, 58(3):321-329, 1974.

Descriptors--Educational Research; *Scientific Enterprise; *Scientific Literacy; *Science Tests; *Tests

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

Purpose

The general purpose of this study was to investigate correlational associations between test instruments which, in the opinion of the investigators, measured "nature of science" objectives. To this end, five research hypotheses were posited. There will be a positive correlation between:

A. The "Nature of Science Scale" (Kimball, 1968) test and subtests of the "Test on the Social Aspects of Science" (Korth, 1969).

B. The "Nature of Science Scale" test and subtests of the "Science Support Scale" (Schwirian, 1968) test,

C. Subtests of the "Test on the Social Aspects of Science" and subtests of the "Science Support Scale" test.

D. The subtests of the "Test on the Social Aspects of Science."

E. The subtests of the "Social Aspects of Science" test.

Rationale

The study is couched in the framework of recent trends in science teaching. That is, a claim is made that science educators have become increasingly concerned that students understand more than facts, concepts, and technological applications of science. Additional concerns have been "that students understand how scientists operate, have correct attitudes about science and scientists, and know what science is really like." Objectives related to these latter concerns are termed "Nature of Science" objectives.

After examining eleven test instruments which "had been constructed and validated for objectives related to the Nature of Science," the investigators decided there was little agreement about the measurement of understanding of nature of science-type objectives even though the instruments "contain many common ideas and statements." The study was undertaken to cast light upon this observation.

Research Design and Procedure

Four criteria were established to select instruments for the study. (1) "The instrument should either be broad and comprehensive in scope or have pertinent subtests." Reasoning—the nature of science as discussed by science educators has various components. (2) "The items on the instrument should be compatible with a Likert-type response format." Reasoning—There is not always a "correct" or "incorrect" position with regard to nature of science objectives. (3) "The instrument should be easily readable by high school students." Reasoning—the study was to be conducted with students in grades 9-12. (4) "The instrument should be relatively short." Reasoning—the instrument could be administered during one class period, and students would not be apprehensive or reluctant to participate if not confronted with a massive, complex instrument.

Three instruments were selected based upon those criteria. The "Nature of Science Scale" (NOSS), the "Test on the Social Aspects of Science" (TSAS), and the "Science Support Scale" (SSS). The TSAS and the SSS had subtests; the NOSS did not.

Using four high schools (grades 9-12), the investigators identified approximately 300 students who were enrolled in science courses. A random one-third of each class was given each instrument.

Findings

Students chose the "agree" response more frequently than the "strongly agree" response for most items and likewise for the "disagree" response and "strongly disagree" response. Before the correlational analysis was undertaken, "strongly agree" and "agree" responses were combined as were "strongly disagree" and "disagree" responses for each item of each instrument.

To test the hypotheses, correlations were tested for significant differences via the _t_ statistic. "(A) Eleven correlation coefficients greater than ±0.21 were considered significantly different from zero" ($\alpha < 0.05$, N = 69). The results are summarized in Table I.

Table I

Significant Intercorrelational Coefficients of TSAS Subtests, SSS Subtests, and NOSS

Variable	TSAS-1	TSAS-2	TSAS-3	TSAS-4	TSAS-5	SSS-1	SSS-2	SSS-3	SSS-4	SSS-5
TSAS-2	.219									
TSAS-3	.294	.329								
TSAS-4	.362		.299							
TSAS-5	HYP.D		.417							
SSS-1						HYP.E				
SSS-2		HYP.C				.282				
SSS-3		-.368				.285	.319			
SSS-4						.249	.324	.490		
SSS-5							.303		.248	
NOSS		HYP.A						HYP.B		

1. Hypothesis A was not supported.
2. Hypothesis B was not supported.
3. Hypothesis C was not supported.
4. Hypothesis D was supported for 6 of the 10 correlations.
5. Hypothesis E was supported for 8 of the 10 correlations.

Interpretations

Possible explanations for students' tendencies to respond "agree" and "disagree" as opposed to "strongly agree" and "strongly disagree" were:

1. reluctance to "go out on a limb";
2. lack of complete familiarity with words and phrases in the instruments' statements; and
3. agreement with part of a complex statement but disagreement with another part of the same statement.

The investigators suggested an interview with subjects as a means of understanding this phenomenon and the underlying causes. However, this possibility was dismissed, stating it "was not a major dimension of the study."

With regard to the resulting tests of hypotheses, the investigators concluded "that the items used to measure broad areas of the nature of science (NOSS) are not related to items that measure pertinent or specific areas of the nature of science (TSAS and SSS)." It was also concluded that the TSAS and the SSS instruments were measuring separate domains within the nature of science since subtests of each instrument were related within but not across instruments.

It was recommended that science educators should give attention to outlining the entire domain of the nature of science in order to attain relationships among the various components. To this end, the investigators developed an outline by elaborating upon the major components.

1. Methods and Aims of Science
2. Characteristics of Scientists
3. Assumptions of Science
4. Processes of Science
5. Interaction of Science with:
   A. Society
   B. Technology

ABSTRACTOR'S ANALYSIS

The reviewer of this research report agrees with the investigators that objectives related to the broad descriptive phrase, nature of science, have been and continue to be important to science educators. There is also agreement that instruments which are commonly construed (not necessarily by the original authors) to assess knowledge or attitudes related to nature of science are very diverse even though they contain "common ideas and statements."

The title and introduction to this research report led this reviewer to hope the investigators had attempted to discover unity in the wild variety of these instruments in a way somewhat akin to Bronowski's description of science as "nothing else than the search to discover unity in the wild variety of nature..." (1965, p. 16). However, the purpose—by now seemingly out of context—stated the intention "to investigate the relationships between three selected instruments..." Translated into research hypotheses, one finds that another correlational study is in the offing.

Given the contradictory nature of the rationale for the study, this reviewer was surprised to see the research hypotheses stated in positive terms. Would it not have been more appropriate to have stated the hypotheses in a neutral manner? That is to say, does not the rationale belie the possibility of expressing directional relationships?

The four criteria and their respective justification for selecting instruments were reasonable. However, they were aimed more at administration of the instruments as opposed to substantive issues related to nature of science objectives. If the study was truly to be "an analysis of several instruments measuring 'Nature of Science' objectives," more criteria should have been offered. An example would be, "The instruments should each contain subtests which purport to assess the same dimensions of the nature of science."

With regard to the study sample, the instruments were administered to "approximately 300 students at four area high schools who were currently enrolled in science courses at grade levels 9-12." In a research report it is important to specify exactly how many subjects were involved. Where were these four high schools located? How did these students compare in terms of: 1) past science courses; 2) interest in science; 3) types of science courses in which currently enrolled; and 4) future aspirations? What was the stratification of students by grade level by high school? A reader is always at a loss to understand and interpret research findings and conclusions if he does not know a great deal about the study population and sample.

With regard to data analyses and the results and conclusions, the investigators stated that, "Prior to the analyses, strongly agree and agree responses were combined as were strongly disagree and disagree responses for each item." They claimed the procedure "is done with most studies using Likert scale responses." No references were cited to attest to the claim or to give credence to the collapse of the data. This reviewer does not recall collapsing of Likert data as a common procedure in the context of the planned statistical analyses. One would wonder why the collapse in response choices was not done prior to administering the instruments if in fact the collapse procedure was planned a priori.

There were three potential explanations offered for the observation that students chose the less extreme responses more frequently than the more extreme "strongly" responses—1) students were afraid to "go out on a limb;" 2) students' lack of complete familiarity with the words or phrases in the statements;" and 3) "agreement with part of a complex statement and disagreement with another part." The latter two reasons raise serious questions about the completeness of the criteria used to choose instruments for the study and the investigators' applications of those criteria which were posited. Additionally, this research report does not describe a pilot study. Based upon the foregoing results and their potential explanations, one is reminded again of the crucial importance of conducting one or more pilot studies prior to a major study.

At this juncture, it is appropriate to point out that no statements were made about instrument reliabilities in the context of this particular study. The investigators evidently have decided to join the ranks of those persons who either do not determine or who do not report reliability values.

The investigators inserted a paragraph into the research report which compared the percentages of responses on the NOSS instrument made by students in their study to those made by scientists, science majors, philosophy majors, and science teachers in Kimball's (1968) study. Not only was it out of context (no rationale or hypotheses prepared the reader for such comparisons), it was inappropriate. An examination of Campbell and Stanley's (1963, pp. 5-6) short monograph will illuminate this criticism.

The hypotheses and their associated statistical tests left this reviewer with several questions. Why were the research hypotheses stated positively when the introduction leads one to believe the investigators found "little agreement about the measurement of understandings" of nature of science objectives? With the research hypotheses stated in a positive sense, why were the correlations "tested for significant differences from zero" using "all correlation coefficients greater than ± 0.21" as being significantly different from zero? The investigators' means of expressing themselves indicate they do not understand differences between research and null hypotheses and differences between one and two-tailed statistical tests.

Additional questions must be raised relative to the statistical tests. Why did the investigators base their tests on a sample size of 69 when there were "approximately 300" subjects in the study? With 11 intercorrelated variables, did the investigators not recognize their resulting decrease in alpha level? Finally, in interpreting their results table, why did the investigators conclude that TSAS-5 was not statistically correlated with TSAS-3?

The investigators ended their report with an outline of their conception of the domain of the nature of science. Without discussing the merits of their outline, this reviewer would only suggest that a research report is not the appropriate medium in which to introduce such a model.

References

Bronowski, Jacob. Science and Human Values. New York: Harper and Row, 1965.

Campbell, Donald T. and Julian C. Stanley. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally College Publishing Company, 1963, pp. 5-6.

Kimball, Merritt. "Understanding the Nature of Science: A Comparison of Scientists and Science Teachers." Journal of Research on Science Teaching, 5:110-119, 1968.

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

Schwirian, Patricia. "On Measuring Attitudes Toward Science." Science Education, 52:172-179, 1968.

Hibbard, K. Michael and Joseph D. Novak, "Audio-Tutorial Elementary School Science Instruction as a Method for Study of Children's Concept Learning: Particulate Nature of Matter." Science Education, 59(4):559-570, 1975.

Descriptors--Audio Active Laboratories; Educational Research; Elementary Education; *Elementary School Science; Instruction; *Science Education; *Scientific Concepts; *Teaching Methods

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Rodney L. Doran, State University of New York at Buffalo.

Purpose

As stated by Hibbard and Novak, "A primary purpose of this study was to understand the cognitive structures which children develop with respect to solids, liquids and air, whether on their own or through formal instruction."

Rationale

Citing research on science concept learning and that utilizing A-T instruction, the authors attempted to "show how pupil achievement of science concepts follows patterns consistent with the learning theory of Ausubel."

A central hypothesis to this subsumption-based learning theory is that "all learning-acquisition of new knowledge or reorganization of existing knowledge is based on new information impinging on the learner, and also on the particular cognitive structure of that learner at the start of the learning tasks."

Research Design and Procedure

The one-shot case study was conducted with "children in four first-grade classes in two elementary schools in Ithaca, New York..." The 84 "instructed" children received a four-lesson introduction to A-T instruction and a "ten-lesson sequence dealing with aspects of the particulate nature of matter and effect of energy exchanges on particles of matter." Each lesson required approximately 18 minutes for completion. The instructional schedule provided for one lesson to be accomplished each week.

The 38 "uninstructed" children received only the four-lesson introductory sequence to familiarize them with A-T methods so they could be tested subsequently via picture tests with audio instructions. The normal first-grade classes had "little or no science instruction, and hence the 'uninstructed' group constituted a control group similar in age and other experiences, but not exposed to A-T science lessons. Students in the 'uninstructed' group were in the same school, but different classrooms, than their 'instructed' peers. No statement of selection procedures was included."

The evaluation procedures used involved two formats:

1. tests including 'production' and 'recognition' questions.
2. 'individual interviews modeled after Piaget's clinical technique.'

An example of a picture-production question was "one which showed a person standing near a clear glass bottle full of 'smelly' liquid. The child was asked to pretend that the smell could be seen and then to draw the smell." A parallel picture-recognition question "showed" a set of drawings, each with the same person and smelly substance but each with a different representation of how a smell might be drawn. The child was asked to select the best representation of a "smell." The production-type questions were administered before the recognition items to avoid cueing. Each child heard the typed instructions to these tests via headphones. Children were tested in groups but they were isolated from each other by cardboard partitions. One hundred picture test questions were organized into four booklets and administered to both groups of students over a two-week period.

Additionally, 25 instructed and 10 uninstructed children were randomly selected and interviewed individually. "The interview technique involved an individual child with hands-on material in a sequence of questions calling for predictions, manipulations, observations, and explanations. Two interview questions were very similar to the picture questions, while the third, called U-tube interview, was described as "a novel problem-solving situation." In this interview coffee is poured into two U-tubes, one containing water and the other only air, and the child is asked to smell at the other end of each tube. Responses to all interviews were taped for subsequent analysis.

The duration of the entire study was apparently 16 weeks (14 weekly lessons and two weeks of picture testing) plus any delay between instruction and testing (which was not specified) and time required for individual interviews.

The children's responses were not scored. "Instead, a system of categories was derived to group responses for each question. The categories were intended to distinguish between students with different cognitive structures in a certain conceptual area, i.e., the relationship between a material and its smell." The authors recognized that there would be some variance in responses among those grouped in any single category. "This grouping of responses, while it may make the data more convenient to manage and may be functional, potentially hides information as to particular differences in children's structures."

Findings

With respect to the responses to the questions requiring students for drawings of smells, the author's descriptions were the following:

1. 66 percent of the instructed (I) and 6 percent of the uninstructed (UI) children used dots to represent a smell. The remainder used lines or some other forms of representation.

2. 30 percent of the I and 2 percent of the UI groups drawing showed the smell going all around the room, not just to a nose. The others drew the smell going just to a nose or one place in the room.

3. 55 percent of the I and 12 percent of the UI drew a model with the smell originating at or within the source of the smell. The other children showed the smell originating above the source.

4. 23 percent of the I and more of the UI group used a dot model which showed the smell coming from the source and going all around the room.

Responses to questions requesting drawings of the "inside of" a solid, a liquid, and air were generally similar to the above, but the authors recognized that some student's drawings are "difficult to interpret and the children's knowledge is underestimated by this type of test." Interviews with selected children were used to probe the limits of cognitive differentiation which may allow some insight into children's learning.

From these results the authors believed that many instructed children seemed to have developed a more sophisticated model for smells and the states of matter—many employing a particulate model. However, they questioned whether "this cognitive differentiation will be useful for problem solving in new situations." Related to this question, the results of the U-tube interview were cited. "The instructed children differed from the uninstructed mainly in their more prevalent use of mechanisms to explain how water could block some of the smell than could the air." Their use of mechanisms may be an indication of the increasing cognitive differentiation in this concept area and ability to, in a qualitative way, manipulate several relevant variables simultaneously."

The students who were classified as using particulate mechanisms to explain the U-tube phenomena (N = 5) also used particulate explanations to the other two interview questions. The children who displayed no mechanism with the U-tube interview were similarly unable to use any explanatory mechanisms in the other interviews.

Interpretations

According to the authors, "this study shows that instructed children use a particulate model to explain the nature of smells much more effectively than uninstructed children, and that they also use somewhat better explanatory models for comparing solids, liquids, and air." Further, some children "demonstrated their mastery of the new models by exploring them to solve a novel problem involving the movement of liquids and air."

The authors concluded that the particulate model used in the instruction "was compatible with the learner's cognitive structure and, therefore, many children were able to synthesize and/or use a particulate model of smells from their previous knowledge and information in the instruction." However, such synthesis did not occur as frequently with respect to explanations of solids, liquids, and air, "presumably because the lesson instruction and the children's prior knowledge were not compatible or the extent of instruction was inadequate."

In summary, the authors inferred that "development of basic science concepts in children's cognitive structure is possible and should result in facilitation of future learning in science and in reading or mathematical skills where these involve science materials."

ABSTRACTOR'S ANALYSIS

This study is not a comparison of results from differing methods of instruction, but an attempt to examine how children's cognitive structures change as a function of relevant instruction. A-T instruction was chosen because of its minimal need of teacher involvement and ease in standardizing instruction. The concepts chosen have been used extensively in past science education research.

While getting any kind of a sample for research spanning several months is difficult, it would seem appropriate to aim for larger samples, especially when, after secondary selection and analysis, one group is reduced to a sample of five. It seems important for the readers to know more about the sample and how it was chosen. Even a few brief comments about the school and the neighborhood would have been helpful. A sample chosen from schools largely composed of children of Cornell faculty may not be representative of U.S. elementary schools. Second, the method for choosing which children were to receive the in-depth instruction should have been specified. Comparisons between cognitive structures of instructed and uninstructed groups as reported here is valid if the groups are shown to be s similar or were chosen randomly from some population. Student self-selection or teacher selection processes have been widely described as potentially having biases. These threats to the validity of a study are important to all styles of research. Assumedly, the first grade level was chosen because it is the beginning year of formal instruction. However, a cross-section of students from other elementary school grade levels would have greatly expanded the nature of possible results and conclusions. The duration of the instructional program seems sufficiently long but more details of the contents of the lessons would have been most helpful.

The authors' review of related research, specifically with respect to science concept learning, appears to be limited. Two of the studies attributed to the University of Wisconsin R & D Center (7 & 9 in original bibliography) were not associated with the UW research program. Further studies from the UW group which had particular relevance to the pictorial dimension of representing particulate models of matters were not cited (e.g. dissertations by Hasan, 1968, and Doran, 1969).

It appears from inspection of the article that the following understandings may be within cognitive structures related to this content domain.

Many of the ideas incorporated here were from the research cited previously by Doran.

Through work such as that done by Hibbard and Novak, it would be possible to support or refute such a model or some part of it. As not all children follow the same learning pattern and therefore possess different cognitive structures, an underlying model or framework would be helpful for instruction and evaluation.

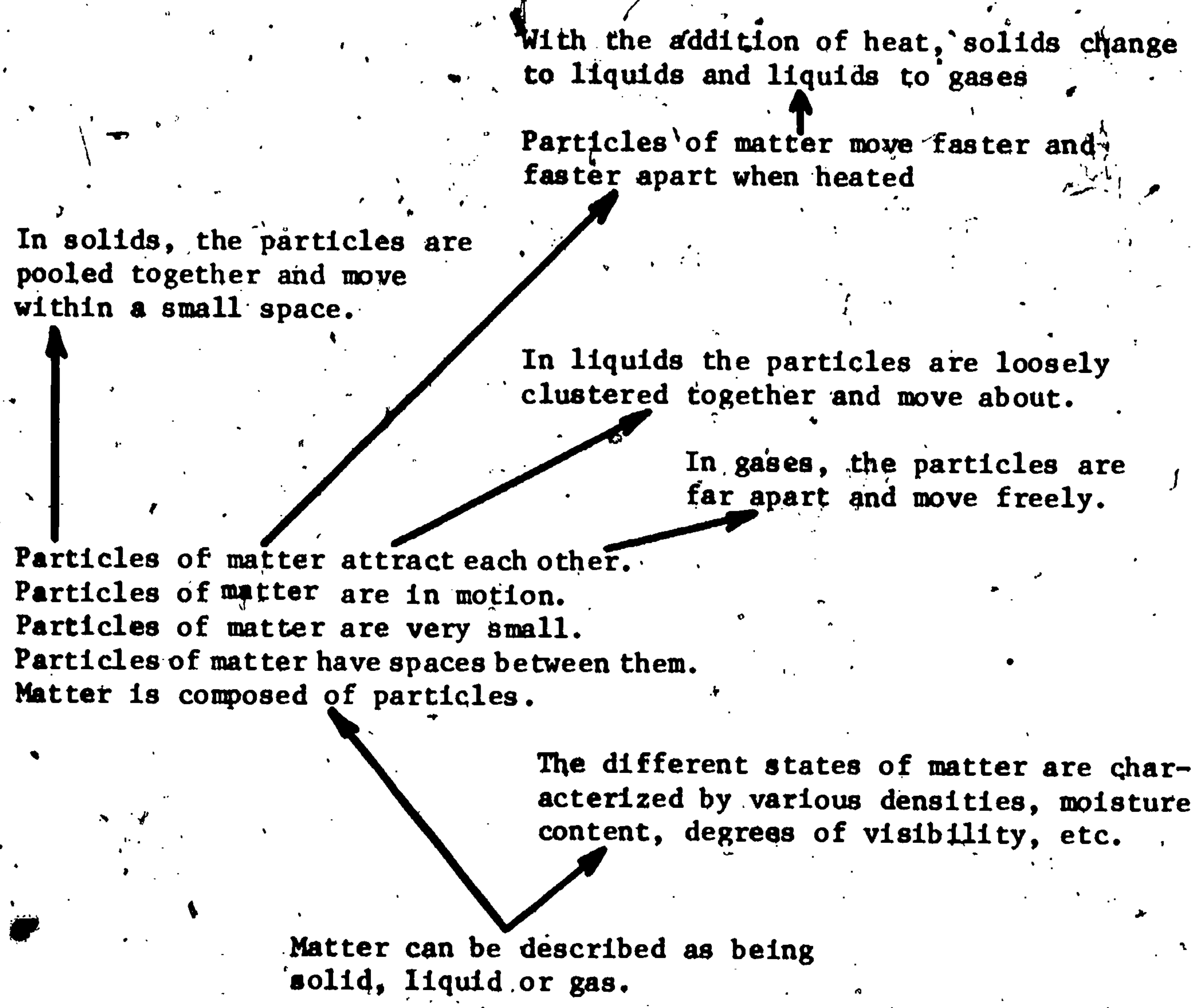

Figure 1.--Hypothesized Structure of Concepts

The testing procedure used in this study was very extensive: 100 questions administered over a two-week period. But no description was included of which phenomena were used in these 100 questions, i.e. how many questions focused on solids, liquids, or gases? Without such information, it is hard to imagine the need for 100 questions. Great care was taken to standardize the instructions and other elements of the data collecting. Results of the "production" questions and individual interviews were discussed in considerable detail, but no mention was made of results from the "recognition" questions.

The categorization of student responses to interviews and their drawings is admittedly difficult. The abstractor is not sure what the investigators uniquely determined from this phase of the study, as contrasted with results from the non-discussed "recognition" type questions. Using questions with distractors constructed to represent specific misconceptions and errors, one can still distinguish between children with different cognitive structure (see Doran, 1972). It seems that with the considerable amount of past research on children's learning of these particulate nature of matter concepts, such would have been possible.

Research in this conceptual area is pivotal because of its use with subsequent instruction on so many topics. It is necessary for parallel studies to explore the ability of the youngsters to manipulate several variables simultaneously and other logical operations described Piagetian theory. These Piaget-based studies could utilize some of the variables such as those used in this study. Item one could begin to describe more fully interactions between cognitive structure (a la Ausubel) and cognitive operations (a la Piaget).

Hibbard and Novak have contributed significantly in a crucial area. It is hoped that they will continue to explore the many dimensions of science concept learning.

REFERENCES

Hasan, Omar H. "A Test Related to the Particle Nature of Matter with A Minimal Verbal Content." Unpublished Ph.D. dissertation, University of Wisconsin, Madison, 1968.

Doran, Rodney L. "Development of Test Items Related to Selected Concepts Within the Scheme The Particle Nature of Matter." Unpublished Ph.D. dissertation, University of Wisconsin, Madison, 1969.

Doran, Rodney L. "Misconceptions of Selected Science Concepts Held by Elementary School Students." _Journal of Research in Science Teaching_, 9(2):127-137, 1972.

Lawson, A. E.; F. H. Nordland; and J. B. Kahle. "Levels of Intellectual Development and Reading Ability in Disadvantaged Students and the Teaching of Science." Science Education, 59(1):113-125, 1975.

Descriptors--*Educationally Disadvantaged; Educational Research; *Intellectual Development; *Learning Theories; Problem Solving; *Reading Ability; Science Education; *Secondary School Science

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Blanche Bante Emerick and Frances Lawrenz, Minneapolis, Minnesota.

Purpose

The purpose of the investigation was to determine the relationship between intellectual development and reading ability.

Rationale

Through references to previous research, the authors establish a logical justification for their investigation. First, a connection between general problem-solving capabilities and reading ability is documented. Then, reading ability is presented as a complex cognitive skill with the suggestion that there is a link between general cognitive ability and reading ability. Therefore, it is suggested that since the Piagetian model is a comprehensive model of intellectual development, a strong positive relationship should exist between reading ability and the level of intellectual development as described by Piaget.

Research Design and Procedure

The investigation examined the correlations among scores on ten Piagetian tasks used to assess levels of concrete and formal reasoning and problem-solving abilities, and scores on the Sequential Test of Educational Progress—Reading Form 3A (STEP-Reading) Educational Testing Service, 1956. The Piagetian tasks were: (1) conservation of number, (2) conservation of substance, (3) conservation of continuous quantity, (4) conservation of length, (5) conservation of area, (6) conservation of weight, (7) conservation of volume using clay, (8) conservation of volume using cylinders, (9) separation of variables, and (10) exclusion of irrelevant variables. Reading scores were obtained for the following: Reproduce Ideas Translate, Make Inferences, Analyze Motivation, Analyze Presentation, Ctiticize, and Total Score.

After administering the Piagetian tasks to 506 randomly selected students and constructing a frequency distribution of their total scores from each of the 10 tasks, a subsample of 35 subjects was selected to complete the STEP reading examination. The subsample was selected to be representative of each of the Piagetian score categories from the frequency distribution. The subjects, 18 males and 17 females, were attending an urban high school which enrolls predominately black and Spanish-American students. The subjects ranged in age from 14.0 to 17.7 years.

Findings

The relationship among the Piagetian scores and the reading ability scores for the 35 students was examined in various ways. A scatter diagram showing the relationship of the Piagetian total score and the STEP total score was presented along with the Pearson product-moment correlation coefficient for the two variables (.70, $p < .001$). An analysis of linear prediction (the hypothesis of no difference between eta^2 and rho^2 was accepted at the .05 level) indicated that a linear and significant relationship existed between the subjects' scores on the two instruments.

The correlations among each Piagetian task and total task score and the reading subscales and total reading score were also presented. In general, the correlations were positive and significant with the more difficult tasks correlating more highly with the reading scales than did the easier tasks. Correlations between the Piagetian total score and each reading subscale were all positive and significant.

Interpretations

The authors discuss why such a strong relationship, as evidenced by their correlational results, should exist between reading ability and the growth of logical thinking ability. Their premise is that before the subject can obtain a high score on the STEP test, he must be able to understand what he has read. Other research on this topic is examined and a tentative conclusion that the development of logico-mathematical mental operations precedes the development of linguistic skills is entertained. Two items from the STEP test are analyzed to ascertain what mental operations are necessary before the questions can be answered. After this analysis, the authors conclude that "To improve reading comprehension..., it may be argued that what must be done is improve a student's thinking ability or increase his level of intellectual development."

After reaching the above conclusion, the authors reassure the reader that the school can indeed affect a student's level of intellectual development, and that an ideal vehicle for this may be through the teaching of inquiry-type laboratory and field activities in science. Three studies are then cited which support this contention.

The authors summarize by stating that their interpretation of the results indicates that the present emphasis on reading instruction as a means of improving language development and intellectual growth is misplaced, and that the provision of situations which optimize the development of the learner's ability to think would prove to be more beneficial. The authors state that "This development is likely to be obtained only when instruction is designed with the present intellectual level of the child in mind and when the child is allowed to interact with the materials of the disciplines, with his peers, and with his instructors, to answer questions which the child himself has raised."

ABSTRACTOR'S ANALYSIS

This review addresses two levels of concern about the article. The more fundamental level characterizes the study as an example of Piagetian research in the behavioralist tradition. To counterbalance that method of research, the constructivist model is introduced. The second level of concern about the article involves the terminology of the article and research procedures used in the study.

Current American research on Piagetian questions falls into two principal categories: behaviorist and constructivist (Emerick and Easley, 1978). One type is exemplified in the article being discussed; namely, a behavioral approach to Piaget's theory of cognitive development. The behaviorist approach postulates a one-to-one correspondence between mental operations and their surface appearance as behavior. On that assumption, the article authors claim that the observance of individual acts is sufficient evidence that certain mental operations are occurring. That type of research emphasizes classifying students based on their performance on Piagetian tasks.

A different type of Piagetian research postulates that the relationships between mental operations and subject behaviors can be described as many-many. Such an assumption means a single mental operation can manifest itself in a variety of behaviors, only a select few of which have been described by Piaget. Similarly, a given behavior can be the result of the functioning of a variety of mental operations. In this second type of research, study focuses on the functioning of those operations and how they relate to observed behavior.

The major proposition presented by the authors of the article is:

> If reading ability can be viewed as being dependent upon, or at least linked to, strategies for attacking and solving problems which themselves are dependent upon the growth of logical thinking abilities, then there should exist a strong positive relationship between reading ability and level of intellectual development as described by Piaget (p. 114).

The authors then describe a study showing a positive correlation between reading ability, as measured by the STEP reading examination, and cognitive development. An analysis of reading test items follows the description, demonstrating the plausibility of the assertion that cognitive development is related to reading test performance. Essentially, the authors contend that high performance on the reading test requires high level reasoning skills such as criticizing and analyzing. Thus, common sense would suggest that improving those reasoning skills would increase performance on such reading tests. On the other hand, improving language ability (which is only a surface indication of mental operations that might be functioning), need not enhance the mental operations generating those surface indicators (behavior).

Given the argument described in the previous paragraph, of what significance are the correlational data which are the heart of the article? Are these data evidence that subjects the hypothesis to a critical test? Isn't a rational demonstration of why the same logical abilities affect behavior on both tests of intellectual development and tests of reading ability more convincing than is a correlational study?

It might be argued that the purpose of collecting data, as in the correlational study, lies in validating theories in "real world" settings. Behaviorists and constructivists part company over what method should be used for such validation efforts. Behaviorists perform statistical tests to compare groups of subjects, while the constructivists concentrate on a small number of subjects in order to construct models of the relationships between mental operations and behaviors.

In the study described in the article, the collection and analysis of data merely confirmed what was later argued in the analysis of the test items, that is, performance on the reading test probably relies on the same logical abilities as performance on the Piagetian tasks. Would it not be more important and interesting to explore the facets of that relationship and its supporting theory?

The remainder of this review will focus specifically on the terminology and sampling procedures in the article.

The arguments in the article are persuasive unless the reader pays close attention to the terminology and the role the terminology plays in the logical development of the argument. Essentially, the terminology is imprecise and, therefore, elusive. Piaget, on the one hand, has defined precisely his meaning of logico-mathematical operations, although the definitions might be difficult to comprehend. On the other hand, in the article, concepts such as problem-solving capabilities, problem-solving strategies, intellectual development, logico-mathematical mental operations, and thinking ability are all terms used without definitions or even regularity. Repetition of terms would be a small price to pay if a few clearly defined terms were to replace the variety of terms currently in the article.

Examples of the elusive terminology are "reading and language ability" and "thinking ability." The authors' argument climaxes with the claim that improving students' thinking ability will improve their reading and language ability. If reading and language ability is defined solely as performance on the STEP test, then such a conclusion is not very interesting, since translating and making inferences, analyzing motivation and presentation, and criticizing certainly qualify as "thinking abilities." In other words, thinking abilities include the reading abilities tested by the STEP test. Thus, the climax of authors' argument can be reduced to improving thinking ability to improve thinking ability. If, on the other hand, the authors wish to argue that reading and language ability is somehow a special form of thinking ability, there might then be some basis for the claim that improving the general skills enhances specific skills. However, the ambiguity of those concepts in the article makes application of the concepts difficult for both educators and researchers.

The methodology of the study raises some questions, especially the Piagetian aspect of the research. An interpretation of Piaget's work different from that in the article contends that Piaget did not describe subjects' behaviors on his tasks so that those behaviors could be used as measures of cognitive development. To limit "correct" responses on a Piagetian task to a single response, such as holding all variables constant except one, is to define mental operations in terms of behaviors rather than operations. Such an approach precludes consideration of subjects' alternative conceptualizations of the Piagetian tasks. Alternative conceptualizations could result in behaviors deviating from those described by Piaget when he discusses particular tasks. For example, if a subject thinks of the variables in a problem as interacting rather than being independent, the subject might test for the effects of combinations of variables instead of the effects of independent variables. The first author has observed such behavior and reasoning in subjects.

The article does not provide a rationale for the point system for scoring subjects on the Piagetian tasks. Readers should be given an acceptable explanation of why a correct response is given credit equal to a correct explanation, or why a III-B classification is worth three times as much as a III-A classification. Awarding points by observing particular isolated behaviors is not a logical extension of Piaget's theory of development in that the "value" of a behavior lies only in the role it plays as intermediary between the subject's mental operations and the environment. Therefore, to investigate the value of a behavior, the researcher must generate a plausible explanation of the interaction of mental operations and environment. Then the explanation should be tested through observation of actual behavior. The hypothesized explanation should be related not only to a particular sequence of behaviors, but also should be based on a set of interconnected postulates (a theory) which attempts to explain a larger piece of reality.

The sampling procedure for the study motivated several questions. Ten Piagetian tasks (only two of which were attributed directly to Piaget) were administered to 506 randomly selected high school students. The subjects then were grouped according to their scores. Out of each group who performed within a range of scores, three or four students constituted the whole group so that random selection was impossible. The practice of combining subsets of subjects who sometimes are randomly selected and sometimes comprise the whole group is a methodologically dubious procedure. Why was the correlation study not done simply on 35 subjects who had been randomly selected from the high school population and who had been administered both the reading test and the Piagetian tasks? And, an n of 35 for this study seems unnecessarily small since it is assumed the reading test is more easily administered than the Piagetian tasks.

Why does the title of the article contain the word "disadvantaged" when no further mention is made of the concept in the article? Is the reader to assume that "an urban high school which enrolls predominately black and Spanish-American students" is by definition "disadvantaged"?

The researchers sampled students in the 14.0 to 17.7 year age range, a span of 3.7 years. The potential correlations between age and stage of intellectual development and age and reading ability was ignored in the article's analysis. Also, it is peculiar that, if the emphasis of the study is on reading instruction, why was it not conducted on younger subjects? It is possible the authors wanted to ensure that both formal operations and high level reading skills were represented in the sample. Assuming that reading skill and intellectual development correlate positively with age, it is fortunate for the researchers that they tested the older subjects. It was at the higher reading skills and intellectual development that the correlation was strongest.

In summary, the authors were ingenious in justifying science teaching in terms of improving reading and language skills. The integration of the two subjects to the mutual benefit of both is a topic worthy of more research. For that topic, it is probably not necessary to incorporate the Piagetian theory of intellectual development. If researchers are interested in the theoretical relationships of reading skills and reasoning skills, then precise definitions and logically developed arguments are minimum requirements for meaningful discussion of the topic.

REFERENCE

Emerick, B. B. and J. A. Easley. "A Constructivist Challenge to the Validity of Tests of Formal Operations." Paper presented at the meeting of the American Educational Research Association, Toronto, 1978.

This paper draws a comparison between the two models of Piagetian research.

Parker, G. E. and Thomas Mertens. "Programmed Instruction, Test Performance, and Classroom Discussion." Journal of College Science Teaching, 4(2):103-106, November 1964.
Descriptors--*Achievement; *Biology; College Science; Classroom Communication; *Educational Research; *Group Discussion; Programmed Instruction; *Programmed Texts; Science Education

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Linda R. DeTure, Winter Park, Florida.

Purpose

The purpose of this study was to investigate the effects of selected programmed units on the test performance and the quality of classroom discussion of students enrolled in an introductory college biology class.

Rationale

Studies examining the effects of programmed instruction on test performance are relatively numerous. In this particular study, the authors have been careful to control spurious factors that frequently influence the results of programmed instruction. The programmed units did not differ from the conventional units in philosophy, emphasis and examples or content. Quizzes were announced well in advance and the Hawthorne effect was controlled by experimental design. The second hypothesis, that students taught by programmed text will engage in higher level classroom discussion, has not been researched as extensively as the first hypothesis concerning test performance. An intent of this study was to add empirical support to the tenet that programmed instruction would enrich classroom interaction, thereby adding a new dimension for the use of programmed instruction as a teaching aid.

Research Design and Procedures

Ninety students were randomly divided into two sections for a second semester introductory biology class. The usual curriculum consisted of 11 units of biology, taught by the lecture-discussion method with each unit having two lectures followed by a discussion and quiz. A test followed a four-unit sequence and a comprehensive final was given. For the study, four programmed units, which were written by the lecturer, were substituted for four conventional units. The four programmed units were Cells, Molecules, DNA and Heredity. Each section, which received programmed instruction for only two of the units, was alternately the experimental or the control group. The design for an experimental sequence consisting of a total of four units could be diagrammed as follows:

$$R_A \quad X_o \quad O_o \quad X_o \quad O_o \quad X_1 \quad O_1 \quad X_o \quad O_2 \quad O_3$$

$$R_B \quad X_o \quad O_o \quad X_o \quad O_o \quad X_o \quad O_1 \quad X_2 \quad O_2 \quad O_3$$

in which X_o = no treatment (conventional class or neutral units) and $X_1 - X_2$ = programmed instruction units, Cell and Molecules, respectively; O_o = unit quiz for the control group, $O_1 - O_2$ = unit quizzes for the treatment group and O_3 = the midterm test.

This design was repeated for the two remaining programmed units. Section A was the experimental group for the DNA unit and Section B, for the Heredity unit.

To examine the test score differences between experimental groups and the control groups, the scores were categorized as high (≥ 80 percent) and low (≤ 79 percent) and subjected to Chi-square analysis. The differences between the classroom discussion of the two groups were measured by two judges who tallied student responses according to an inventory of high-low discussion level indicators. The judges were also asked to write an evaluation of the discussion and to note any difference between the two groups. They were not told which groups were program instructed. The discussion results were also analyzed with Chi-square analysis.

Findings

The Chi-square results indicated that a significantly greater number of students fell into the high category than into the low category for three of the four programmed units. No significant differences were noted for the DNA unit. On the two midterm tests and the final exam the subjects continued to score significantly higher on the program-taught material than on the conventionally taught material.

Differences between the two groups were less striking for the classroom discussions than for the test performance. The only unit in which a statistically significant difference occurred between the two groups was the unit on Molecules. However, the groups who were program taught scored slightly higher in the discussion analyses for every unit, and when the tallies from the four units were cumulated, a statistically significant overall difference occurred between the two groups.

Interpretation

The findings of the report indicated students taught with the programmed units improved in both test performance and classroom discussion. Due to the experimental design, which controlled for the Hawthorne effect, differences in content, emphasis, examples and the use of unannounced quizzes, the author attributed the changes in performance solely to the variation in instructional techniques. Also the final exam scores demonstrated that the initial difference between experimental and control group did not fade by the end of the semester.

On a course evaluation form students acknowledged that they learned more from the programmed units than from the conventional units. However, they also indicated that the lectures were more interesting and enjoyable. They checked two programs as the desirable number of programmed units to be supplemented into the course. The results of this study suggest that programmed textbook supplements could be used as an effective teaching aid to bring meaningful variety into the classroom.

ABSTRACTOR'S ANALYSIS

This study varied from other studies of programmed instruction in some critical areas. One important one is that the content of the lectures and programs were parallel as the lecturer had written the programs. An advantage of any programmed instruction is that it provides immediate feedback and reinforcement of response. This seems to be any effective procedure initially, but students often become "program wise" and learn to make the correct response without really having to process the information. A criticism of programmed instruction is that differences tend to fade with time. The author mentioned this and attempted to measure it by giving a comprehensive final at the end of the course.

However, I think the notion of differences fading with time refers to the effectiveness on the program being in continuous use. If initial interest is high due to the novelty of the teaching procedure, this could account for differences in performance. The differences may disappear or fade with time after the students become accustomed to the programmed method.

On the evaluation form the students requested that no more than two of the eleven units be taught with programmed texts. The researchers suggest that the programs be utilized as supplements to the conventional procedures rather than as a complete substitution for them. In this context programs could serve as an effective instrument for adding variety to normal classroom procedures.

The experimental design was effective for measuring the objectives and it helped control threats to internal validity that frequently plague studies of programmed learning. The authors attempted to avoid the Hawthorne effect by having each section alternately serve as the experimental group and by making intragroup comparisons. In effect this could have served to heighten the novelty effect described by Bracht and Glass (1968), especially since each class was exposed to only two programmed units out of a total eleven. The uniqueness of the method possibly resulted in the students focusing their attention more carefully on the content of the unit. If the programs are utilized to supplement and to add meaningful variety as the authors intended in this study, then it may be worthwhile to trade on the effects of novelty as long as it is kept in mind that the program is not uniquely responsible for the differences in performance.

One question that needs to be raised in the study is the appropriateness of using Chi-square analysis to test the differences between the groups. The Chi-square is usually used for nominal data or the median test is used to measure central tendencies of the two groups if analysis of variance would result in violating the assumptions, which does not seem to be the case here. No information was given concerning the range of the scores or how they clustered within the groups. If the standard deviation was low and if the mean was near the division point between groups, it would appear that the numbers in the two groups could be manipulated to meet the standards for significance. It would have been interesting to see the means and standard deviations and the F values of the groups tested with an analysis of variance. A rationale for the decision to use Chi-square analysis rather than analysis of variance would help clarify these questions.

Overall the study was written in a style that was easy to follow. By using the programmed text on a limited basis, the researchers demonstrated how effective a programmed unit can be in improving test performance and classroom discussion and they avoided the problem of boredom that often accompanies a programmed course. The hypothesis that programmed texts used on a limited basis results in better test performance is supported. The results of the impact of programmed instruction or classroom discussion are somewhat supported, but the strength of that relationship is questionable.

REFERENCES

Bracht, G. H. and G. Glass. "The External Validity of Experiments." American Education Research Journal, 5(4):437-467, 1968.

Stanley, J. C. and D. T. Campbell. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally, 1963.

Young, R. K. and D. J. Veldman. Introductory Statistics for the Behavioral Sciences. New York: Holt, Rinehart and Winston, 1972.

Scott, Norval C., Jr. "Cognitive Style and Inquiry Strategy: A Five-Year Study." Journal of Research in Science Teaching, 10:323-330, 1973.

Descriptors--*Cognitive Ability; *Cognitive Measurement; Educational Research; Elementary School Science; *Inquiry Training; *Longitudinal Studies; Science Education; Secondary School Students; *Teaching Methods

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Gerald G. Neufeld, Brandon University.

Purpose

The study investigated the long-term effects of a two- to three-year exposure to an inquiry teaching strategy on students' cognitive styles. The cognitive styles of the experimental and comparison subjects were assessed one and six years after the end of the treatments.

Rationale

In recent years researchers have become interested in the tactics individuals use to perceive and organize external stimuli. These tactics are known as cognitive styles. One aspect of the research into cognitive styles and their educational implications has focused on the type of cues individuals use in perceiving similarities and, subsequently, in categorizing their environments. This has led to the definition of several "styles of categorization" and their measurement through the use of tests such as the Sigel Cognitive Style Test employed in this study.

Kagan, Moss and Sigel (1963) have found that individuals show a steady, continuous progression from a relatively global style of categorization toward a more analytical one. In addition, they found that an individual's style is resistant to large-scale change because individuals generally retain their relative stylistic positions relative to each other.

Several investigators have attempted to shift students' styles toward more analytical ones. Davis (1967) reports little success in training students to be more analytical in concept identification tasks. Scott (1966, 1970) reported some success but his study did not attempt to assess whether the treatment had any lasting effects. This study attempted to assess the permanency of style changes in students.

Research Design and Procedure

The paper presents the results of two separate but related studies. The author refers to one as a longitudinal study and the other as a cross-sectional study.

The independent variable for both studies was the level of exposure to the inquiry strategy teaching method originally proposed by Suchman (1960) and modified by Scott (1964). Two levels of treatment were used: no exposure and two or three years exposure during the subjects' upper elementary and junior high school years. In the longitudinal study the exposure occurred in science classes during grades 4, 5 and 6 while in the cross-sectional study it occurred during grades 5, 6 and 7. There is no mention of the actual amount of instructional time spent using the inquiry strategy nor any indication of procedures used to ensure that the strategy was implemented appropriately. However, the author indicates that careful questioning of the subjects revealed that the comparison subjects had never been exposed to this strategy and that the exposure of the experimental subjects was limited to the treatment period.

The dependent variable for the studies was the students' styles of categorization as measured by the Sigel Cognitive Style Test (SCST). The SCST consists of a series of cards with three pictures on each card. For each card the subject indicates which pairs of pictures are related and give reasons for the choices. Each grouping and the reasons given for that grouping are scored as belonging to one of six styles of categorization: descriptive-part whole (analytical), descriptive-whole (non-analytical), relational-contextual, categorical functional, categorical-classnaming or categorical-attribute. A student's total score for a category is simply the number of times he or she has grouped pictures using that style of categorization. For this study only the twelve (12) picture cards common to the M (male) and F (female) forms of the SCST were used. The split-half reliability for this modified form of the SCST was found to be 0.76 (n = 101).

The subjects for the studies were drawn from four large, urban high schools. For the longitudinal study there were 16 experimental subjects and 16 comparison subjects. The comparison subjects were drawn from an area which, in 1966 (when the first test was administered), had a similar socioeconomic and cultural background to that of the area in which the experimental subjects lived and had similar tenth grade achievement test scores. The subjects for the cross-sectional study consisted of 26 experimental and 34 comparisons. The comparisons consisted of two classes of high ability high school students. Not all of the experimentals were in the high ability track—thus the comparison group had a slight edge over the experimental group academically. There is no indication of random assignment of students to treatments—the students appear to have selected to receive the treatment.

The research design used for the longitudinal groups is shown diagramatically in Figure 1. The design appears to be a modification of the static-group comparison design described by Campbell and Stanley (1963).

n = 16	X	O_1	O_2	X - occurred in grades 4, 5 and 6
n = 16		O_1	O_2	O_1 - at the end of grade 7
				O_2 - prior to high school graduation

Figure 1: Experimental Design for the Longitudinal Study

The research design used for the cross-sectional groups is shown in Figure 2. It is a static-group comparison design (Campbell and Stanley, 1963) with the post-treatment observation delayed for several years after the treatment. The author implies the experimental subjects for this study were the same students that took part in his previous study (Scott, 1970). However, he does not repeat the previous data nor indicate how these students' scores changed over time—the only comparisons made are with the comparison group.

n = 26	X	0	X - occurred in grades 5, 6 and 7
n = 34		0	0 - prior to high school graduation

Figure 2: Experimental Design for the Cross-sectional Study

Statistical comparisons of the experimental and comparison groups were made for each of the six sub-scores of the SCST for each of the three testing periods (two for the longitudinal study and one for the cross-sectional study). Chi-square was used to test for significant differences between the experimental and comparison groups. A 2 x 2 contingency table was constructed for each of the six SCST categories with the median score in each category being used to determine whether an individual's score was high or low. The four cells were low experimental, high experimental, low comparison and high comparison.

Findings

The analysis of the results of the longitudinal study indicated that students who had been exposed to the inquiry teaching strategy had significantly higher descriptive-part whole (analytical) scores on the SCST than the comparison groups for both the 1966 testing and the 1971 testing. There were no significant differences between the longitudinal groups for any of the other five SCST categories.

The analysis of the cross-sectional results indicated that the experimental group had significantly higher scores on three SCST categories: descriptive-part whole, descriptive-whole, and categorical-attribute, than the comparison group.

The median scores of the two comparison groups (the longitudinal and the cross-sectional) appeared to be quite similar for all SCST categories except for categorical-classnaming. The author attributes the apparently higher performance of the cross-sectional group on this category to the fact that all the students in this group had high verbal abilities and were in the top academic track in school whereas not all the members of the longitudinal group were in this track.

A non-statistical comparison of the two experimental groups on the pre-graduation tests indicated that the median scores for the longitudinal group were consistently lower than those of the cross-sectional group for all categories of the SCST. The author attributes this to the fact that the cross-sectional group had more hours of exposure to the inquiry method.

Interpretations

The author concludes that, since similar results were obtained from both the cross-sectional and longitudinal studies, the inquiry strategy method influences a student's ability to classify objects towards a more descriptive-part whole (or analytical) style. He cites a number of studies to support his view that the development of an analytical ability has important implications for increasing student achievement in such disciplines as mathematics and science. However, he indicates that the development of such an analytical orientation could hamper creativity and productivity in such areas as the arts and humanities and points out the need for further research to assess the total impact of inquiry strategies on students.

ABSTRACTOR'S ANALYSIS

Although there is a long history of research in education and psychology relating to how individuals perceive their environments, process sensory information, form concepts, and solve problems, it is only within the last 25 years that research in these areas has been organized and integrated using the concept of cognitive style. Much of the research on cognitive styles has focused on the field-dependent-field-independent dimension. The research relating to this dimension and its educational implications are summarized in a recent review (Witkin and others, 1977). Kagan, Moss, and Sigel (1963) defined a related dimension which they termed a person's style of categorization. The Sigel Cognitive Style Test (SCST) was developed to assess the various styles of categorization. This conceptualization of styles and the SCST do not appear to have been widely accepted by the research community and are cited only rarely in the recent literature.

The study under review appears to build logically on the conceptualization by Kagan, Moss, and Sigel (1963) and related studies (Davis, 1967 and Scott, 1967 and 1970). However, there are few follow-up studies since this area of research appears to have fallen into disfavor. Another reason for a lack of follow-up studies may be the recent emphasis on adapting instruction to the learner's aptitudes and cognitive styles rather than attempting to modify the learner's aptitude or cognitive styles to suit the discipline. The more recent approach is more defensible on moral and ethical grounds because modifying something as basic as an individual's cognitive style may have undesirable side effects. The possibility of a detrimental side effect of a shift to a more analytical style is indicated by Lee, Kagan and Rabson's (1963) finding that an analytic set could hamper creativity and productivity in areas such as the arts and humanities.

The validity of the study is weakened by the research design employed. The static-group comparison design fails to control for selection and mortality as well as for interactions between selection and such factors as maturation and the experimental treatment (Campbell and Stanley, 1963). The study would have been improved by the use of a research design such as the one depicted in Figure 3 which combines the elements of a time series design and a nonequivalent control group design to increase internal and external validity. The design would have been further strengthened if the subjects had been randomly assigned to treatments. However, random assignment is not usually possible in the real world of classroom research.

$$\frac{O_1 \quad X \quad O_2 \quad O_3 \quad O_4 \quad O_5}{O_1 \qquad\quad O_2 \quad O_3 \quad O_4 \quad O_5}$$

Figure 3: An Improved Research Design

The reader would have more confidence in the validity of the findings if the author would have been more specific about the extent and nature of the experimental treatment. The author indicates that the experimental subjects were exposed to the inquiry strategy over a two or three-year period but provides no indication of the actual number of hours spent using this method. The author appears to have this information because he states that "the cross-sectional group had more hours of exposure to the inquiry method. . .than the longitudinal inquiry students" (p. 329). Unfortunately, the author did not include this information in the report. There is also no indication that the classrooms were monitored to ensure that the intended treatment occurred as planned.

It would be unfair to criticize the author for using the relatively rare SCST because when his research began in 1962 there were no other measures of cognitive style suitable for use in classroom situations. However, this does point up the danger of attempting research, especially longitudinal research, in an area where standard tests have not been developed.

Since the SCST is not well known, the author should have been more careful in describing both the test and the scoring procedure used. The text of the report appears to indicate that the SCST is an ipsative test—the sum of the six sub-scores for each subject equals twelve because each subject responds to twelve cards with each response being assigned to one of the six categories. However, a study of the medians given in the data tables and a related report (Scott, 1970) indicates that the SCST is only semi-ipsative since a subject may make more than one response per card. Greater clarity in this section of the report would have been helpful in interpreting the results.

The author describes the procedures he used for making the statistical comparisons with refreshing clarity. However, he neglects to mention that the procedure he is following is for the median test for two independent samples (Siegel, 1956). The procedure he followed is correct, provided $n_1 + n_2$ is greater than 20 and the expected frequency in every cell is greater than 5. Hopefully, these conditions were met. A reader would have more confidence in the results of the tests if the author had named the test—the reader would then assume that the author was at least aware of the limitations on the tests.

The position of the author's statement that the rejection level for the statistical tests was set at 0.05 implies that he recognized the need for establishing rejection limits before the tests were run. The author deserves commendation and emulation for this fact alone. It would be a welcome change if the educational research reflected the fact that researchers recognized the need for establishing rejection limits before the fact and then used estimated effect size and power analysis (Cohen, 1971) to determine the sample size they needed.

REFERENCES

Campbell, D. T. and J. C. Stanley. "Experimental and Quasi-experimental Designs for Research on Teaching." in Gage, N. L. (ed.) Handbook of Research on Teaching. Chicago: Rand McNally & Co., 1963.

Cohen, J. Statistical Power Analysis for the Behavioral Sciences. New York: Academic Press, 1969.

Davis, J. Concept Identification as a Function of Cognitive Style, Complexity and Training Procedures: Technical Report 32. Madison: Research and Development Center for Cognitive Learning, University of Wisconsin, 1967.

Kagan, J.; H. Moss; and I. Sigel. "Psychological Significance of Styles of Conceptualization." in Wright, J. C. and J. Kagan (eds.) Basic Cognitive Processes in Children. Monograph of the Society for Research in Child Development, Vol. 28, No. 2. Lafayette, IN: Child Development Publications, 1963.

Lee, L.; J. Kagan; and A. Rabson. "The Influence of a Preference for Analytical Categorization upon Concept Acquisition." Child Development, 34:433-442, 1963.

Scott, N. "Science Concept Achievement and Cognitive Functions." Journal of Research in Science Teaching, 2:7-16, 1964.

Scott, N. "Strategy of Inquiry and Styles of Categorization." Journal of Research in Science Teaching, 4:143-154, 1966.

Scott, N. "Strategy of Inquiry and Styles of Categorization: A Three-year Exploratory Study." Journal of Research in Science Teaching, 7:95-102, 1970.

Suchman, J. "Inquiry training in the Elementary School." The Science Teacher, 27:42-47, 1960.

Siegel, S. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill, 1956.

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):1-64, 1977.

Vanderbroucke, A. C., Jr. "Evaluation of a Personalized Instruction System in General Chemistry." Journal of Chemical Education, 52(8):516-518, 1975.

Descriptors--*Autoinstructional Programs; *Chemistry; *College Science, *Course Evaluation; Educational Research; Higher Education; *Instruction; Programmed Instruction; Science Education

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Robert W. Johnson, Towson State University.

Purpose

The purpose of the project was to explore an alternative to the lecture format widely used in general college chemistry. The model selected for field testing in this study was the Personalized System of Instruction, PSI, developed by F. Keller.

Rationale

Recent attempts to improve student performance in college science have resulted in new approaches to individualizing instruction. Programmed texts, modular self-pacing units of study, and computer-assisted instructional formats are being used more widely in the higher education. Encouraged by the results reported by other investigators using the self-pacing mode, the author decided to test the PSI model against the more traditional lecture format.

Research Design and Procedures

The study sample was drawn from eight sections of general chemistry (111 and 112) during the school years of 1972-73 and 1973-74. Students knowingly selected either A sections (lecture/mornings) or B sections (PSI/afternoons). Students, mainly freshmen and sophomores, were pursuing programs of study in biology and the health-related sciences.

Both treatment groups shared common (1) textbooks, (2) weekly laboratory periods, and (3) final examinations. A two-hour department final exam concluded the first semester of study. The ACS - Cooperative Examination in General Chemistry was given at the end of the second semester.

Students in Section A were given four 50-minute lectures per week. Section B students were given one 50-minute lecture per week plus three 50-minute testing periods per week. Additionally, B students were aided by paid tutorial upperclass chemistry majors. They monitored student progress through 20 units of material demanding 90 percent mastery of each unit. Comparison data for evaluating the two different instructional formats were gathered from scores obtained on both final examinations. Correlation co-efficients between the five ACT (American College Testing) parameters and three variables were obtained for the eight sections. The three variables examined were (1) final examination scores, (2) total points earned in course, and (3) cumulative grade point averages.

Findings

Based upon a comparison of final test scores at the end of the first term of study, students in both A and B sections demonstrated about equal performance. Similar results were attained by both groups completing the second term of study in 1973-74. However, for the second term of 1972-73, PSI students achieved significantly higher scores on the ACS - Cooperative Examination than did their lectured counterparts.

Of the five ACT Test parameters, student mathematical scores were found to correlate with each of the three variables more frequently and with the highest level of confidence.

Reported also was the fact that a far higher percentage of A and B course grades were awarded to PSI students over the two-year trial period.

Interpretations

The author concluded from the test results that PSI students were learning at about equivalent levels to students of comparable ability in lecture sections. Nonetheless, the mean score of the 1972-73 PSI sections were higher on the ACS - Cooperative Examination, the second semester chemistry final exam. This difference in performance was attributed to better mathematics ability demonstrated by PSI students on the ACT Test and not to the method of instruction.

In evaluating the course, pupil observation supported the notion that PSI students left the course with a better feeling about chemistry. The author suggested in his assessment of the course that PSI students appeared more actively involved in the course and attained better retention of material. He implied that self-pacing developed in students a more positive feeling about chemistry and that, in general, a better attitude about learning was engendered.

ABSTRACTOR'S ANALYSIS

The study, herein reviewed, is but one of many being reported in the literature. It attempts to compare and evaluate student performance resulting from two different approaches to college instruction, e.g., lecture versus self-pacing study. The findings here fail to support the superiority of one instructional format over the other. Nevertheless, results do underscore the contention that the PSI approach can be an effective alternative teaching/learning model when certain conditions are met. Teachers do teach and students do learn. But does it necessarily follow, then, that all students learn better when taught by the self-pacing method? The larger body of research purporting to compare the effectiveness of PSI to the lecture approach seems to suggest an answer in the negative. One generalization is most frequently offered by investigators assessing the Keller model: PSI enhances and improves student attitudes toward learning. The author of this study concurs with that assessment.

The Personalized System of Instruction, admittedly, is a fairly recent innovation at the college level. The advantage is that students, given precise instructional objectives, can move through a module of material and master it at their own pace. Its major limitation resides within its own methodology. The habit of procrastination may not be effectively managed by some self-pacing students.

Therefore, to suggest that only PSI students become actively involved in the learning process is to discount the concept of individualizing instruction. Learning styles differ greatly among students. Teaching styles and formats also differ. The problem, then, is to find a proper fit between teaching format and a particular learning style which produces maximum student performance. Certainly, PSI may be one of many approaches to effective instruction.

Given the requisite tools, skills, and cognitive development, the idea of making a student personally responsible for his/her learning is not a new concept in education. It was not the purpose of this study to advance or explore new ground.

The study's research design was carefully planned and executed. The sample was limited but adequate, and controlled for comparison purposes. Quantitative performance data in the form of student test scores were collected and tabulated in table form. Correlation coefficients between selected parameters and significant variables were obtained at the 0.01 and 0.001 confidence limits. Comparative analysis of data gathered from the two treatment groups produces no significant evidence to suggest that a Personalized System of Instruction was better or worse than lecturing to students in college chemistry courses.

One wonders if other variables aside from instructional format might not have a greater effect upon improving student performance. Individualized learning styles of students most certainly affect the teaching/learning relationship just as teaching style does. Instructor enthusiasm for his subject or a dynamic personality may just preclude the superiority of one instructional approach to another. Gifted teachers can be found amongst lecturers, demonstrators, and facilitators in education throughout history. What attributes are the benchmarks of gifted and effective teachers? What formal training in theory and practice should be required to produce competent teachers in higher education? These questions pose problems for on-going research. What this study does support strongly is the contention that PSI options should be made available to students in courses that are freely entered into by competent faculty.

Perhaps the best way to insure improved student performance is to guarantee improved instruction. Much still needs to be done in the area of training college teachers to teach effectively. Irrespective of degrees held, many college teachers have little or no professional background in education. None hold professional licenses or certificates to practice their teaching skills in higher education within a single state. Maybe state licensing boards need to be established to set standards and review credentials of those seeking teaching positions in higher education. None presently exist.

Reference

Keller, F. S. "Good-bye, Teacher . . ." Journal of Applied Behavioral Analysis, 1:79-90, 1968.

RESPONSES TO ANALYSIS

A RESPONSE TO THE ANALYSIS OF

Moore, Kenneth D. and Jacob W. Blankenship. "Teaching Basic Science Skills through Realistic Science Experiences in the Elementary School," by Donald E. Riechard for Investigations in Science Education, 5(4):12-18, 1979.

by

Kenneth D. Moore
University of Science and Arts of Oklahoma

In reviewing the abstractor's analysis of our article, several comments appear to be in order.

First, the use of the word "frugal" in the leading sentence of the second paragraph of the Written Report section appears to be misleading as to the abstractor's intention. The use of this word suggests that the abstractor found the study lacking with respect to our presentation of related research and with respect to a rationale for the study. However, this is inconsistent with the subsequent comments in this paragraph.

The abstractor further notes that the study is weak with regard to validity and reliability. Although "construct validity" of the assessment instrument has been established and reported, the abstractor is correct in reporting that concurrent or predictive validity of the assessment instrument has not been established. This limitation should be kept in mind by potential users of the study results. Since it was not expedient to administer the assessment instrument to each respondent a second time, I tend to agree that reliability has not been rigorously established. However, it should be noted that this research study deals with averages and these averages will, of course, be much more reliable than individual responses. Moreover, the randomization procedure utilized in this study greatly enhances the reliability and external validity of the findings. The authors therefore feel the results can be generalized to the elementary teacher population with a fair degree of confidence.

In his closing remarks, the abstractor concludes that the study does not create any new or unique knowledge. He further suggests that "providing realistic science experiences" becomes a high priority need factor only after other higher priority need factors have been met. These closing remarks suggest that the abstractor incorrectly interpreted the findings.

The authors suggest that the identified factors (Table 2, page 340, Science Education article) as well as the high loading items of each of the identified high priority factors needs to be reexamined. Although the authors do agree that this study confirms the findings of past research in that the most appropriate programs for science education appear to be those which provide "realistic science experiences," the study goes beyond mere confirmation. The results suggest that science teachers view "providing realistic science experiences" as being a higher priority need than such needs as those in the areas of discipline, reading, writing, mathematics, and spelling. Furthermore, an examination of the high loading items on factor need I "Providing realistic science experiences" suggests some of the constraints which seem to be keeping science teachers from providing these more appropriate "realistic science experiences" in their classrooms.

While there is disagreement on several points, the abstractor has made several enlightening and valid points which, if addressed, would make this study and future similar research much stronger. We welcome and wish to thank him for his critical analysis of our work.

A RESPONSE TO THE ANALYSIS OF

Peterson, Kenneth D. "Scientific Inquiry Training for High School Students: Experimental Evaluation of a Model Program," by John R. Staver. Investigations in Science Education, 5(4): 29-36, 1979.

by

Kenneth D. Peterson
University of Utah

I appreciate the critique of my recent research report (Peterson, 1978) which appeared in Investigations in Science Education (Staver, 1979). It is of service to researchers to have pertinent and responsive reviews.

There were a number of questions raised in the critique which, I think, are answered in the original article:

1. Only one scientific inquiry assessment instrument was used in the study. It was administered to half of each of the three treatment groups as a pretest and to all subjects as a posttest. No pretest experience effects were found.

2. Variance explained by treatment (omega squared) for each of the 15 dependent variables ranged from .02 to .64 (not .04).

3. The scientific inquiry training treatment group (SI) was superior to the verbal instruction treatment group (VI) on 6 of the 15 variables; the VI group was superior to the SI group on none.

The posttest means and standard deviations are presented in Table 1. They were omitted because of space limitations.

Table 1

Posttest Results--Treatment Scores: Means and Standard Deviations

	TREATMENT GROUP					
	Project Physics (N=24)		Verbal Instruction (N=17)		Science Inquiry (N=26)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
1. Number of Variables	4.29	1.51	8.65	3.35	9.77	3.53
2. Variable Points of View	1.71	.53	2.47	.82	2.96	.68
3. Uncued Variables	.12	.26	1.00	.92	1.12	1.20
4. Divergent Variables	.21	.42	1.06	.89	1.96	1.94
5. Number of Questions	4.17	1.57	8.06	2.54	7.12	2.63
6. Question Points of View	1.83	.55	2.88	.84	3.12	.85
7. Uncued Questions	.37	.56	.59	.93	1.00	1.55
8. Divergent Questions	.54	.69	.76	.71	1.85	1.93
9. Question Criteria	.96	.75	1.88	1.52	2.77	.95
10. Experimental Components	1.79	1.13	2.53	1.56	4.19	1.06
11. Number of Generalizations	1.25	1.13	1.53	.98	1.27	.50
12. Form of Generalizations	.71	.43	.53	.53	.65	.48
13. Additions to Investigation	2.08	1.02	2.53	1.42	2.15	1.41
14. Science Processes	4.33	2.72	4.65	1.94	8.23	2.15
15. Relations Among Processes	.42	.49	.65	.40	.96	.14

My reading of Ausubel suggests that it was reasonable to expect high school seniors to be at an "abstract-verbal" stage and not at the "concrete-intuitive" stage which it replaces in "the latter portion of the junior high school period" (Ausubel, 1963, pp. 133-5). Thus, it was anticipated that the verbal instruction group, which received instruction as per Ausubel, would have performed as well or better than subjects completing other types of instruction. (The VI instruction was designed to "teach to test" to the same degree as the SI instruction).

The reviewer discounts my conclusion that the Ausubelian predictions did not hold for some kinds of scientific inquiry performances. He offered the interpretation that the students were at various stages of Piagetian development (i.e., some not yet at a level of formal operations). While the literature cited certainly backs this point of view, his statement changes Ausubel's explanation. The result is a synthetic interpretation which combines both Ausubel and Piaget. While this makes sense to me, and is probably a more adequate theoretical explanation, it is not pure Ausubel. One point of the study was to test Ausubelian theory. I still offer the conclusion that Ausubelian principles do not hold uniformly well for all kinds of scientific inquiry learnings.

REFERENCES

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

Peterson, K. D. "Scientific Inquiry Training for High School Students: Experimental Evaluation of a Model Program." Journal of Research in Science Teaching, 15(2): 153-159, 1978.

Staver, J. R. Expanded Abstract and Analysis of "Scientific Inquiry Training Program for High School Students: Experimental Evaluation of a Model Program." Investigations in Science Education, 5(4): 29-36, 1979.

A RESPONSE TO THE ANALYSIS OF

Suter, Patricia E. "Using Audio-Visual Study Lessons to Teach the Underprepared Student," by Elizabeth Kean. Investigations in Science Education, 5(4): 37-41 , 1979.

by

Patricia H. Suter
Del Mar College

Your inclusion of my article "Using Audio-Visual Study Lessons to Teach the Unprepared Student," published in School Science and Mathematics, is greatly appreciated. I would like to make the following comments on the critique:

1. The course of instruction for which the lessons were designed is now what we classify as "introductory." The college has two other courses using this designation, which are taken by liberal arts students or health science students. The course taken by students using the lessons is called General Inorganic Chemistry which is taken by those expecting to attain a bachelor's degree in a field of science, mathematics, or engineering.

2. The term "better student" refers to those receiving grades of "A" or "B". The final examination given in the course is that prepared by the American Chemical Society and these students compare well with the national norms.

3. The lessons were prepared to help students learn the required material. Instructors at Del Mar College operate under the theory that it is their job to tell the student what he or she must learn and to provide all possible help in that learning process. The A-V Study Lessons have provided some of that help. Over 300 students a year use one or more of the lessons. The students themselves are the judge of the effectiveness of the lessons and they continue to report that they find them to be of great assistance. Therefore, we continue to believe that the preparation of these lessons was worth the time, effort and

money expended in their preparation. We are continuing to modify the lessons as the need arises.

4. The background of the students attending Del Mar College is diverse. Many have language difficulties as well as problems in dealing with mathematical concepts. Entrance examination scores continue to decline nationwide in both these areas. Our students are attending college to become better able to compete in a technological, English speaking work situation. No bilingual lessons were prepared, nor are contemplated. Our Spanish speaking students are not more adept in that language than in English because most learn a mixture of English and Spanish at home. Other languages (German, Czech, Chinese, Vietnamese) are the first languages of a significant part of the student population. No student has ever requested the lessons in some other language, even though anonymous polls are taken each semester and a suggestion box is available for the students to use. The lessons are modified periodically, either to incorporate student suggestion, or to include new material.

5. Follow-up studies of the students registered in General Inorganic Chemistry have shown that most (81 percent) of the students who failed or dropped the course did not use the study lessons at all. Many of these students are working and have little time for study. The usage of the study lessons increases during the latter part of the first semester and the second semester. Probably this is due to unsuccessful students dropping out or failing, and the remaining students finding that the lessons do help them learn. It may also be due in part to some students who had high school chemistry realizing that they do not know as much chemistry as they thought they did at the beginning of the first semester. Some have commented that this is so.

6. The members of the chemistry faculty at Del Mar College are not trained in educational theory. We are merely trying to provide aids for our students to assist their learning of the required material. They are given detailed lists of objectives for each unit of

study and glossaries of new technical terms. No other control of the vocabulary is attempted on the tapes nor in the classroom. However, unusual words are defined the first time they are used.

7. Design details are available from the author. (Write: Patricia H. Suter, Associate Professor, Dept. of Chemistry, Division of Arts and Sciences, Del Mar College, Corpus Christi, Texas 78404.)

8. Presently the faculty at Del Mar College is offering other study aids to our students. These are CAI units, some multiple testing after the Keller plan, and supervised tutoring. The students are not required to use any of these aids. The responsibility for learning is, in our opinion, the student's. We are there to help and encourage the process.