DOCUMENT RESUME

ED 164 326 SE 025 938

AUTHOR	Helgeson, Stanley L., Ed.; Blosser, Patricia E., Ed.
TITLE	Investigations in Science Education, Vol. 3, No. 4. Expanded Abstracts and Critical Analyses of Recent Research.
INSTITUTION	Ohio State Univ., Columbus. Center for Science and Mathematics Education.
PUB DATE	77
NOTE	67p.
AVAILABLE FROM	Information Reference Center (ERIC/IRC), The Ohio State University, 1200 Chambers Rd., 3rd Floor, Columbus, OH 43212 (Subscription $6.00, $1.75 single copy)
EDRS PRICE	MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS	*Abstracts; Autoinstructional Methods; Biological Sciences; *Cognitive Development; *Educational Research; *Instruction; Research; Research Methodology; *Science Education; Student Attitudes; Values
IDENTIFIERS	*Piaget (Jean); Research Reports

ABSTRACT

This issue provides abstracts and analyses, prepared by science educators, of research reports in three areas of investigation. The first area is cognitive development. Three of the five studies presented in this area are Piagetian in nature. The second group contains four studies dealing with attitudes and values. And the third area, auto-tutorial instruction, contains two studies dealing with biological sciences, one at the high school and one at the university level. Each abstract includes bibliographical data, research design and procedure, purpose, and research rationale, along with the abstractor's analysis. (BB)

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, Ohio 43210

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

TABLE OF CONTENTS

Volume 3, Number 4, 1977

TABLE OF CONTENTS (Continued)

NOTES . . .

from the Editor

This issue of INVESTIGATIONS IN SCIENCE EDUCATION contains three clusters of research reports. The first cluster, COGNITIVE DEVELOPMENT, reflects a continued emphasis on Piagetian research. Three of the five studies in this group are Piagetian in nature.

The second cluster ontains four studies dealing with ATTITUDES AND VALUES, with two of these studies being reviewed in a single critique.

The final cluster, AUTO-TUTORIAL INSTRUCTION, contains two studies both dealing with biological sciences, one at the high school and one at the university level.

We continue to encourage publishable responses to the analyses and invite suggestions for improvement of INVESTIGATIONS IN SCIENCE EDUCATION.

Stanley L. Helgeson
Editor

Patricia E. Blosser
Associate Editor

COGNITIVE DEVELOPMENT

Bredderman, Ted. "Elementary School Science Experience and the Ability to Combine and Control Variables." Science Education, 58(4):457-469, 1974.

Descriptors--*Cognitive Development; Curriculum; Educational Research; *Elementary School Science; Laboratory Experiments; *Learning Theories; Science Education; *Scientific Methodology; Science Course Improvement Project

Expanded abstract and analysis prepared especially for I.S.E. by Anton E. Lawson, Arizona State University.

Purpose

This study attempted to answer three questions with regard to the ability to combine and control variables:

1. At what grade levels do children from upper middle class American communities develop these abilities?

2. Does the elementary school science program in the school they attend produce a noticeable effect in this development?

3. Is there a relationship between a child's ability to control variables and his ability to form combinations of variables?

Rationale

A trend in recent years has been noted in which elementary science programs such as Science--A Process Approach have begun to emphasize the investigative processes of the scientist. These investigative processes involve the abilities to combine and control variables. These abilities have received widespread attention in the science education and psychology literature in recent years. This interest can be traced to Inhelder and Piaget's 1958 volume The Growth of Logical Thinking from Childhood to Adolescence in which the combination and control of variables play a central role in the developing "formal operational" reasoning of the adolescent.

In Inhelder and Piaget's theory of intellectual development, formal operational reasoning represents the fourth and final stage of development. Formal operational reasoning, which can be considered hypothetico-deductive "scientific reasoning" is theoretically dependent upon the prior development of concrete operational reasoning of the child. Thus formal operational reasoning is delayed in appearance in most persons until adolescence. Further, Inhelder and Piaget theorize that the development of the "combinatorial system" which allows the generation of all possible combinations of variables is necessary for the subject to devise controlled experiments and to properly deduce their implications.

Research Design and Procedure

Design. The research design employed in this study is referred to by Campbell and Stanley (1963) as the one-shot case study design. Subjects were taught science in one of three programs: (1) a noncoordinated science program, (2) a specialist-run laboratory program, and (3) the Science—A Process Approach program. Following science instruction in one of these three programs subjects were posttested with measures of their ability to combine variables and to control variables. IQ scores, measured by the Otis Quick Scoring Mental Ability Test, were taken from school records.

Subjects. Subjects were 240 children in grades 4, 6, 8, and 10. A random sample of 20 subjects was chosen from each of the three programs at each grade.

Procedure. The ability to combine variables and to control variables was measured using individual tasks administered by four trained examiners.

Tasks. The combination of variables task consisted in presenting the subject with pieces of wooden men consisting of four different pieces varying in color of shoes, trousers, jacket and hat. In all, 16 different men could be constructed. Subjects were asked to generate as many wooden men as they could. Responses were scored into one of seven categories reflecting the degree to which their approach was systematic and complete.

The control of variables task consisted of presenting the subject with two identical lever systems in which five independent variables affected the

distance the lever moved when a weight was placed at its end. Questions were posed in which the subject was asked to test the effect of the independent variables. A numerical score reflecting the number of variables deliberately controlled was calculated.

Data Analysis. Subjects' abilities to combine and control variables were compared with respect to grade level and elementary school science programs using Chi-square and analysis of covariance techniques. Correlation coefficients (type not specified) were computed and reported among task responses, sex, grade level, age, IQ, and mental age.

Findings

The ability to systematically generate combinations of variables increased significantly with age ($p < 0.001$) up to the eighth grade then leveled off. The ability to control variables also increased significantly with age ($p < 0.01$). This increase was evident between each age group with no leveling off between the eighth and tenth grades.

No significant differences were observed among the three different science programs on either task ($p < 0.05$).

The ability to systematically combine variables was found to be significantly related to the ability to control variables ($p < 0.001$). Nearly all subjects who could combine variables could also control variables. However, those who could combine variables may or may not have been able to control variables. This relationship was reversed when only subjects with a highly developed ability to control variables were considered or when all subjects were considered and their control of variables responses were divided into non-perfect and perfect categories.

The correlation analysis revealed moderate to high intercorrelations among the abilities to combine and control variables, grade level, age, IQ, and mental age. By and large, sex did not correlate significantly with task performance.

Interpretations

The following conclusions were drawn:

1. There is a significant improvement in children's abilities to combine and control variables during the pre- and early adolescent years.
2. These improvements do not seem to be noticeably affected by the science program used during the upper elementary grades.
3. The initial development of the control of variables ability precedes that of combining variables but final mastery is not achieved until the combinatorial ability has been acquired.

The development of the abilities of combining and controlling variables appears gradual and spans several years. Further, at any one age the variability in performance is considerable.

On both tasks some fourth graders performed better than some tenth graders. The failure of subjects from the Science—A Process Approach program to perform better than the subjects from the other programs on the combination and control of variables tasks was unexpected. This failure was attributed to lack of transfer of learning.

The third conclusion is in quite close agreement with theoretical expectations based upon Piagetian theory which states that the experimental isolation of variables leads to the combinatorial system which enables one to use propositional reasoning which in turn allows complete mastery of the control of variables procedure.

ABSTRACTOR'S ANALYSIS

This research addresses fundamentally important issues in education; i.e., what reasoning processes develop or fail to develop with age? And what can instruction (in this case science instruction) do to facilitate this development?

The finding that components of formal reasoning do develop with age and that a wide range of performance exists at any one age has been corroborated by a number of more recent studies (e.g., Lawson, Karplus and Adi, in press; Martorano, 1977; Moshman, 1977).

The issue of determining sequences of development of components of formal reasoning, in this case the development of the ability to combine variables prior to the development of the ability to control variables, is a very difficult issue to deal with adequately. Such analyses are complicated by the fact that logically equivalent tasks can vary in degree of difficulty due to logically irrelevant variables (e.g., Nummedal and Collea, 1978) thus making the question of which logical ability comes first difficult to determine. One would need to insure that variables such as task content and complexity were controlled. This was not the case in the Bredderman study. Recall the combinations item involved wooden men and colors of shoes, trousers, etc. while the control of variables task involved levers. It would have been better to have asked the subjects to generate combinations of the variables involved in the lever task as well as control of variables questions. Likewise the subjects could have been asked control of variables questions with the wooden men as well as the combinations questions. Nevertheless, the tasks that were used appear to be novel to the subjects and the variables involved were familiar enough to be understood so that the results may in fact not have been affected by the difference in context. See Lawson (1978) for a more complete discussion of this issue.

The conclusion that the Science—A Process Approach (S-APA) program did not noticeably advance subjects' abilities to combine and control variables is of course regrettable. On the other hand, Linn and Thier (1975) reported that the Science Curriculum Improvement Study (SCIS) program did have a noticeable effect on advancing subjects' ability to control variables. Failure in one instance and success in the other could be due to several factors, the programs themselves being just one possibility. Other possibilities include the adequacy with which each program was taught, the length of time devoted to each program, and of course to the method of measuring student reasoning abilities.

If, in fact, the improvement in one instance and lack of noticeable improvement in the other can be attributed to the superiority of the SCIS program, this could be due to the fact that the SCIS program makes every attempt to embed the logical reasoning within the context of the science concepts being taught whereas the S-APA program does not. For example, SCIS class lessons are designed to study environmental factors, energy sources, photosynthesis and so on, with the combination and control of variables coming into play only when needed to understand such science concepts.

The S-APA program, on the other hand, has lessons in which students explicitly study controlling variables, observing, measuring and the like, with the concepts of science relegated to a secondary role. This difference may be significant in that the SCIS program's approach may be more in tune with the way logical reasoning "spontaneously" manifests itself with advancing age. Children are typically not directly taught to combine and/or control variables outside of school yet the evidence clearly shows that those logical abilities to develop (at least in many children).

I believe that Bredderman's conjecture is essentially correct that the reason the S-APA program subjects failed to outperform the other subjects was due to a failure to transfer their learning. The Lawson and Wollman (1976) study clearly indicated that fifth and seventh grade students of normal intelligence can acquire the ability to control variables in a relatively short period of time. But do they use this ability once it is acquired? Does it transfer to new situations?

The answer to these questions, I believe, is simply that it depends. First of all, it depends upon the way in which the ability was taught. It must be taught in such a way that it is connected to the students' own intuitions (intuitively connected). And it must be taught in such a way that it is related to many external contexts, not just science contexts (contextually connected). Further, it depends upon the sort of environment in which the student finds himself after the training. If he returns to an environment in which the newly acquired ability is not needed (i.e., an authoritarian environment in which questions are not raised and evidence is not sought to test tentative answers to these questions), the ability will atrophy from disuse. However, if he returns

to an environment in which the newly acquired reasoning ability is demanded, it will be further sharpened and refined. Transfer will be effected.

The successful program then must be more than a program in reasoning in science. The program must begin with examples from the students' own experience and it must continue with examples outside of the context of science. Further it may be necessary that many courses (not just the science course) create situations in which answers to questions are actively sought and tested.

The Bredderman study is a valid and worthwhile initial attempt to determine the effectiveness of a science program on the development of reasoning of students. It clearly points out the magnitude of the task at hand.

An extremely important direction for future research is in testing various hypotheses concerning methods of teaching logical reasoning and variables affecting the transfer of such teaching.

REFERENCES

Bredderman, T. "Elementary School Science Experience and the Ability to Combine and Control Variables." Science Education, 58(4):457-469, 1974.

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

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

Lawson, A. E. "Expanded Abstract and Critical Analysis of Quasi-Simplex Analysis of Piaget's Operative Structures and Stages." Investigations in Science Education. (In press.)

Lawson, A. E.; R. Karplus; and H. Adi. "The Acquisition of Propositional Logic and Formal Operational Schemata During the Secondary School Years." Journal of Research in Science Teaching. (In press.)

Lawson, A. E. and W. T. Wollman. "Encouraging the Transition from Concrete to Formal Cognitive Functioning: An Experiment." Journal of Research in Science Teaching, 13(5):413-430, 1976.

Linn, M. C. and H. D. Thier. "The Effect of Experimental Science on the Development of Logical Thinking in Children." Journal of Research in Science Teaching, 12(1):49-62, 1975.

Martorano, S. C. "A Developmental Analysis of Performance on Piaget's Formal Operations Tasks." Developmental Psychology, 13(6):666-672, 1977.

Moshman, D. "Consolidation and Stage Formation in the Emergency of Formal Operations." Developmental Psychology, 13(2):95-100, 1977.

Nummedal, S. and F. Collea. "Factors Affecting College Student Performance on Proportional Reasoning Problems." Paper presented at the National Association for Research in Science Teaching Annual Convention, Toronto, March, 1978.

Rowell, J. A. and P. J. Hoffman. "Group Tests for Distinguishing Formal from Concrete Thinkers." Journal of Research in Science Teaching, 12(2):157-164, 1975.

Descriptors--*Cognitive Development; *Developmental Tasks; Educational Research; *Group Tests; Science Education; Secondary Education; *Secondary School Science.

Expanded abstract and analysis prepared especially for I.S.E. by Jerry G. Horn, Kansas State University.

Purpose

The principal objective of the research was to produce workable group forms of Piagetian tasks. The report is a description of an attempt to translate two Piagetian-type problem situations into forms suitable for administration to groups, together with marking schemes, and a report and discussion of the results of their experimental trial.

Rationale

In 1966 Bruner maintained that mental growth was essentially discontinuous and therefore best described by stage developmental theories rather than theories postulating gradual accretion processes. The stage development viewpoint, particularly in the form proposed by Piaget, has gained a steadily increasing importance in the thinking of educators. A number of attempts have been made to construct school science curricula fitted to developmental levels and to analyze existing courses in the same terms. Examples of these attempts are the "Australian Science Education Project" and, in Britain, the "Schools Council 5/13 Project." Ingle and Shayer, in a series of articles, have described a technique for assessing science courses based on Piaget's developmental stages, and have demonstrated its usefulness by providing a close analysis in these terms of the Nuffield O-level courses in chemistry and physics.

While attempts to adapt curricula to the developmental levels of the intended population must be applauded as an important step in improving the quality of our educational system, the problem still remains for the teacher—how to identify the mental development of each child who faces him. Shayer has suggested the indirect method of utilizing mental age

as a predictor of stage of mental development. Since individually administered type examples, as discussed by Inhelder and Piaget, are prohibitively time consuming for use in the normal classroom situation, this study is an attempt to translate two Piagetian-type problem situations into forms suitable for administration to groups.

Research Design and Procedure

The principal objective of the research reported here was to produce workable group forms of Piagetian tasks. The tasks selected for this study were (1) The Chemical Color Change Problem and (2) The Pendulum Problem. The researchers wanted a chemical task and a physical one that were structurally similar but which involved very different subject matter. This was desired in order to test the hypothesis that a particular quality of thought demonstrated in the subject area tends to be demonstrated with problems possessing a similar structure in another subject area.

Essentially, the students composing the sample for the study were given a set of materials, instructions and work sheets and they were asked to complete the task as described in the instructions. There were two work sheets following the initial page of instructions for each task. The first work sheet was designed for the student to record his physical manipulations of the apparatus. The second work sheet asked for a summary of what had been found out in the experiment.

The sample for this study consisted of a total of 193 students (110 boys and 83 girls) who participated in the chemical experiment; 189 of the same students (107 boys and 82 girls) participated in the pendulum experiment. The students were selected from a South Australian metropolitan high school and represented each of the first four years of the high school with the exception of the first year, where there is a random assignment to classes. Students at the school are streamed by ability; therefore, a wide range of abilities was insured by selecting sections at various levels.

The scheme for analyzing the results was, insofar as possible, in accordance with the schemes proposed by Inhelder and Piaget. This involved

placing children in two subcategories, A and B, of each of the two major categories of mental development being considered—concrete (II) and formal (III).

The inter-judge reliabilities obtained from categories given independently by each of the two authors and expressed as product moment correlation coefficients (r) were: chemical test r = 0.96, SE = 0.006 and pendulum test r = 0.92, SE = 0.011.

Findings

The data were presented as frequency and percent of subjects' responses falling into two levels each of the concrete and formal developmental levels. Data from both problems clearly showed the dual trends of increase in percentage of formal thinkers with increase in chronological age and the higher percentage of formal thinkers in the upper stream at the various grade levels. The degree to which the two problems provided the same measure of developmental sublevel, expressed as a product moment correlation coefficient, was r = 0.56; S.E. = 0.05.

Interpretations

According to the researchers, the major conclusion which emerges from this study is the fact that it does seem possible to translate into group form, administer, and assess rapidly with considerable reliability Piagetian-type problem indicators of development level. Of secondary importance and provided only as comments by the researchers, the results suggest that (1) the pendulum problem provided an earlier indication of the onset of formal thinking than did the chemical one, and (2) although having carried out the pendulum experiment only two weeks before the testing reported here in a teaching situation, a substantial number of students from second and third year classes of intermediate ability level gave concrete operational level answers to the problem. This brings into question what is achieved by the attempted teaching of content which, in developmental terms, is beyond the grasp of those performing the exercise.

ABSTRACTOR'S ANALYSIS

The purpose and the procedures used in this investigation are rather unique. This uniqueness complicated the efforts of this abstractor to construct an expanded abstract for Investigations in Science Education, as per the suggested format.

The researchers identified a very practical problem and have attempted to provide a product or at least a model of a product for practitioners of several professions, teachers, curriculum developers, curriculum planners, etc.

There has certainly been an attempt to base this study on the conclusions and writings of others, such as Piaget, Inhelder, Shayer, Ingle and others. To this abstractor's knowledge, there have not been substantial efforts to develop group measures of developmental stages using manipulative equipment and materials. Perhaps some of the recent studies of Robert Karplus are directly applicable to this needed area of investigation.

Essentially, this study showed that the two problems dealing with chemical change and pendulums will result in high school students' responding at various levels of sophistication. The researchers implied that this is an indication that the two problems identify an individual's developmental level. The validity of this inference is based on the fact that older students provided higher level responses. Since individually administered tasks are widely accepted, the validity of this study and the implications of this study would be greatly enhanced if there had been attempts made to compare the same individual on group and individual assessments.

There is difficulty in generalizing the results of this study to a large population, as the sample was selected from one South Australian high school and there was not a detailed description of the sample provided in the report. For example, the subjects were identified as being "streamed" by ability, but there was not any further explanation in this regard.

The research design used cannot be easily identified with any described by Campbell and Stanley. The procedures were really designed to validate

a procedure and the report is a description of that effort. The authors, Rowell and Hoffman, reported on a procedure that is somewhat removed from the classical research designs. The report itself could have provided a more systematic explanation of the steps and clearly have shown where inferences are made. From a reader's standpoint, the differentiation of fact and inference is difficult to determine. While some conclusions may seem logical, they are in fact generalizations that are not evident, at least from the information provided by the authors in this report.

In a practical sense, the authors have provided a ray of hope for a teacher who may be faced with the enormous task of matching curriculum to stages of development. The prospect that an individual's stage of development possibly may be determined in a group setting is encouraging. This study has approached the problem and could have planted the seed for more expansive and generalizable studies.

Renner, John W. and Anton E. Lawson. "Intellectual Development in Preservice Elementary School Teachers: An Evaluation." Journal of College Science Teaching, 5(2):89-92, 1975.

Descriptors--College Science; Curriculum Development; *Educational Research; Higher Education; Instruction; *Intellectual Development; *Learning Theories; Science Education

Expanded abstract and analysis prepared especially for ISE by Robert E. Yager, University of Iowa.

Purpose

The question investigated was whether or not instruction (materials and procedures) can be shown to promote formal reasoning ability in students. The central hypothesis tested was that "concrete curriculum materials" and "inquiry-oriented procedures" would promote gains in levels of intellectual functioning. The study was designed to illustrate the effect of a specific kind of schooling upon the acquisition of formal reasoning abilities.

Rationale

The study arises from Piagetian philosophy. In fact, a quote from Piaget is used as support for the theme:

> ...education is not simply a contribution that would be superimposed on top of the results of an individual development regulated in some inborn way, or that is accomplished by the family alone. From birth to the end of adolescence, education is one whole, and is one of two fundamental necessary factors for intellectual and moral formation, so much so the school carries a great responsibility regarding the final success or failure of the individual in pursuit of his own potential and adaptation to social living.

The authors review recent "evidence" that suggests that 40-60 percent of the U.S. adult population do not acquire abstract reasoning ability. They cite some of their own work at Purdue University which suggests that significant numbers of college students can operate only with concrete ideas and materials. They suggest that this failure to acquire abstract reasoning ability may result from the use of "inappropriate strategies and materials" at the senior high and collegiate levels. The authors

advance the notion that concrete thinkers can be confronted with concrete problems and that "meaningful inquiries" can facilitate formal reasoning ability in students.

Research Design and Procedure

Twenty students were selected from 300 students enrolled in a biology course for elementary teachers as the experimental group. Seventeen other students were selected for the control group. Pretests were administered to all students at the end of their freshman year. The tests were Piagetian tasks that had been used in previous studies concerning all 300 students enrolled. The experimental group was enrolled in "the undergraduate Pre-service Teacher Education Program" while the control group was enrolled in "a traditional physics course for elementary school teachers." After one year (two semesters of instruction) posttests consisting of the same battery of tasks were administered to both groups.

After the pretests were administered, all subjects were categorized into substages of intellectual development. The substages included: 1) early concrete operational, 2) middle concrete operational, 3) late concrete operational, 4) post-concrete operational, 5) early formal operational, 6) middle formal operational, 7) late formal operational.

The six Piagetian tasks used were: 1) conservation of weight, 2) conservation of volume using clay, 3) conservation of volume using metal cylinders, 4) separation of variables, 5) exclusion of irrelevant variables, 6) equilibrium in the balance. Two "trained examiners" conducted the 30-minute interviews. The pretests were completed in April of 1973 and the posttests April of 1974. The interviewers had no knowledge of whether the subjects examined were in the control or experimental groups. The scoring procedures and testing materials were described in a previous publication. The tasks were given to each subject in the same sequence. "Level of subject responses was evaluated on the basis of the difficulty of the task, the quality of the student's verbal responses, and his ability to exhibit the appropriate behaviors; i.e., to exhibit conservation reasoning, to control variables on the exclusion and separation of variables tasks, and to demonstrate understanding of proportional reasoning on the equilibrium task."

Pretest and posttest scores for each of the 37 subjects (20 in the experimental and 17 in the control) were recorded and analyzed.

Findings

The mean gain in substage level was significantly greater for the experimental group. The Mann-Whitney U Test was used to test the differences. In the case of the pretest mean level there was no significant difference. Neither were there significant differences for the posttest mean levels. Hence, the significant mean gain in level for the experimental group is the major finding of the study.

Seven of the 20 subjects in the experimental group gained two substages and 10 gained one substage. Four of the control group subjects showed a pre- to posttest gain of two substages while one showed a gain of one substage. Nine subjects (from both groups) showed no change and three showed a loss of one substage.

Since the mean pretest score for the control group was high, the possibility of a ceiling effect was checked by an analysis of covariance. The F-ratio (removing the effects of the pretest) revealed no significant difference at the .05 level.

The evaluators were also compared. There were no significant results between evaluators on pretest scores, posttest scores, or mean gain scores.

Interpretation

The authors feel that the results support the concept that the entire notion of what constitutes a worthwhile education would change if the primary goal of educators was to be the production of persons who can think in the abstract. They recommend that content be viewed with respect to two questions: 1) Does the content selected accurately portray the discipline? Is the content representative of the discipline? 2) Will a particular segment of content be useful in assisting an adolescent or young adult in using formal reasoning abilities? There would therefore be a new basis for

selecting content. The authors also argue that the results suggest the need for greater involvement of students with the materials of the discipline. They argue further that the results substantiate Piaget's long-standing axiom that "social transmission is an important and extremely necessary component of learning and movement toward the acquisition of formal thought."

ABSTRATOR'S ANALYSIS

It must be kept in mind that this research report has appeared in a publication that is not addressed to the researcher. Hence, many of the criticisms may be explained by noting the intended reader of the article. Nonetheless, some of the generalizations and some of the problems with the design should be considered. Erroneous conclusions can often result from over generalizations and unanswered questions concerning design. Perhaps there would be fewer generalizations and more information available in a research report prepared primarily for active researchers. However, one has only the brief manuscript for purposes of critical analysis.

-Researchers need to exercise more care in categorizing students (and people in general) as to their level of intellectual development. Certainly there is some arbitrary definition with respect to the seven substages described. However, one needs to realize that the categorization may refer to one or more tasks. Much valuable information may be lost when the researcher uses a total score from a variety of tasks. Further, some erroneous conclusions may result. The researchers could (and perhaps should) report the effects of the experimental design upon each task rather than blend the results together for a single total score. The researchers are to be commended for the use of the individual interview and for the careful comparison of evaluator performance.

Relating the study to curricula, teacher educators, and the growing information from Piagetian studies is a commendable feature. Certainly the attention given to such research in both curriculum design and teaching methodology suggests the importance of the study. Its appearance in a journal for the non-researcher suggests its interest and importance as well.

In view of the major problem with duplicating such experiments at other centers, more information concerning the exact protocols would be helpful. The researchers can be excused for their omission in view of the publication and the audience. However, much more precision in terms of describing specific experimental procedures is necessary before the experiment could be repeated independently. The researchers also assume that the tasks do measure the behaviors and/or the levels as described by Piaget. The literature continues to suggest research problems and unanswered questions in these areas.

Some questions concerning the use of parametric statistics instead of non-parametrics can be raised. Again, the careful researcher is anxious to be free of possible criticism concerning the choice of statistics and/or being able to meet all conditions for procedures chosen.

Finally, the researchers use many terms and phrases as if they have universally accepted definitions. Perhaps more precision in writing is even more important for the practitioner than for the researcher. At least the researcher is likely to request more adequate information before accepting results and interpretation than is the practitioner. Some examples of this concern are: What is an "inquiry approach?" How does one recognize "concrete curriculum materials?" What characterizes a "laboratory?" Are all so-called laboratories places where inquiry occurs? What is "teaching through inquiry?" What characterizes an "integrated science experience?" What makes an examiner "trained?"

Some questions are raised by reference to instructors "providing" experiences and students "taking" courses. There seems to be some internal inconsistency concerning philosophy of the researchers (or instructors) and the descriptions used—at least those other than the special terms that are not adequately defined.

The researchers fail to establish that the experimental and control groups are identical. Were all 300 students in the freshman group? How did some get into Biology 205? Were some taken out of the class and enrolled in physics? How does enrollment in the undergraduate Preservice Teacher Education Program (the experimental group) differ from enrollment in a

traditional physics course for elementary school teachers? Were the students in the experimental group selected for that course randomly? Were the 17 in the control group also selected randomly? More description of the courses and the strategies would be desirable. Was the physics group (control) self-selected?

The article is basically a very good report. It is clearly written and one that is a contribution to the field. It represents well the state of research in the area of Piagetian-type studies. There are unanswered questions, some problems of design and interpretation, some disagreement. But that is why it is so important and so valuable as an area of study.

Reports such as this one are badly needed. Too often Piagetian studies merely verify what has been done or report differences from study to study or culture to culture. This article is a fine attempt to use the information to affect programs. Much more research and effort is now needed to bring exciting research information to the position of making a difference in education. The authors end with a timely charge: the time has come for college and university to examine critically what they are teaching in individual courses, why it is taught, and how it is being taught. This reviewer applauds the charge and this research effort!

Linn, M. C., and H. C. Thier. "The Effect of Experiential Science on Development of Logical Thinking in Children." Journal of Research in Science Teaching, 12:49-62, 1975.

Descriptors--*Curriculum Evaluation; Elementary Education; *Elementary School Science; Learning Theories; *Logical Thinking; Science Education; Science Course Improvement Project

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

Purpose

Linn and Thier (1975) conducted an *ex post facto* study (Kerlinger, 1964) to assess the effect of selected units from SCIS on the development of logical thinking in fifth grade children.

Research Design and Procedure

The design chosen was basically a nonrandomized posttest-only control group design. The major independent variable was the study of at least the Energy Sources Unit from the SCIS program. The criterion variable was mean score on Ss' ability to explain in writing selected phenomena (inclined planes) dealing with compensating variables that were presented to the students during the experiment.

The Ss were enrolled in intact classrooms of fifth and eighth grade students; the fifth grade students either had or had not completed Energy Sources, but the eighth graders had not. In the experimental task Ss were shown film clips of the motion of a cart on an inclined plane under different (compensating) conditions of starting height and surface friction. The main criterion variable was assessed through written free response to a question designed to elicit an explanation of the phenomenon which was observed.

Findings

It was concluded that SCIS students understand compensating variables better than did non-SCIS students in urban fringe, suburban, and rural areas of the country.

Interpretations

The study also reviewed logical thought according to Piaget and the rationale for the use of SCIS in developing and studying formal thought. More specifically, it was argued that understanding compensations involving two known variables is an indication that the student has reached the end of the concrete stage and ready to begin the formal stage of development.

ABSTRACTOR'S ANALYSIS

Although the authors used the words experiment and experimenter, the reviewer must classify this study as *ex post facto* or correlational in nature. The latter form of research is by no means unimportant for the bulk of much useful research (e.g., research linking smoking and forms of cancer in human beings) is by and large correlational. Correlational research, however, does allow for certain errors of interpretation and this reviewer will attempt to identify some of those potential errors.

The researchers should be commended for their study relating theory and curriculum development using large samples from varying types of school districts (suburban, urban fringe, and rural) and many states. The study provides a good example of a nationwide data collection and analysis effort conducted at a distance and in an unobtrusive manner. However, some alternative plausible hypotheses (threats to internal validity of the design) are present in the study; it is hoped that their examination will serve to further the cause of experimental research on effects of science instruction.

The key problem is that, as much as we would like to, we are unable to draw the causal inference from this study that selected SCIS experiences cause understanding of compensating variables or logical thinking in the population. Self- and other selection processes, potentially present in *ex post facto* research, strongly suggest the phenomenon that whenever innovations such as national science curricula are adopted; better teachers are selected (and trained), better students are chosen, and extensive materials are purchased. One might infer that most curricula would "succeed" under such conditions. For example, implementation of SCIS programs is usually accompanied by some vendor training or in some other cases, training by science educators. The link between inservice training and degree of implementation has been firmly established in the literature (e.g., Ashley and Butts, 1971; Crowther, 1972). Self-selection and concomitant associates such as inservice training are seen as one source of differential selection bias in this study. Another source is the selection of subjects, who were presumably identified by school personnel. A significant alternative plausible hypothesis is that improved performance in the sample is an artifact of differences between the treatment groups which are unrelated to the treatment.

It can be argued that students with more articulate written skills will outperform students with less articulate skills on a written explanation. Indeed, general ability is posed by the researchers to explain observed sex differences (p. 59). One might similarly attribute higher mean scores of the SCIS fifth graders to superior general ability of "better" students selected for SCIS. No indication is given that general ability was measured or controlled to test this alternative hypothesis. To the extent that the above argument is valid the causal nature of SCIS on the criterion is severely questioned. It also calls in to question the researchers' assertion "...it follows that experimental teachers were generally effective in teaching..." (p. 59).

A further problem common to curriculum studies deals with implementation. While many studies rely on reported use as the measure of implementation (Persall, 1972), such reports may reflect only an attitude of acceptance rather than the possession of necessary knowledge and skills. Fullan and Pomfret (1977) reviewed studies that suggest a positive relation

between degree of implementation and achievement scores. They observe that without teacher implementation behavior changes, student process outcomes will probably not occur. The degree of implementation of Energy Sources in this study seems to be an unknown quantity.

The reviewer is open to an apparently valid criticism for using subsequent research in analyzing the earlier work of Linn and Thier. The only defense pleaded is that the reviewer's goal is to improve research skills which enable us to acquire valid knowledge; it is not to criticize.

Recommendations

In light of the above arguments, it would seem that a series of experimental studies with certain characteristics would better test the causal relation between curricula and formal thinking. First, the rationale between process or formal thought skills targeted in the curricula and those reflective of the acquisition of formal thought should be clearly established. Next, a design should be constructed in which students and teachers are randomly assigned to treatment or control, with variations in general ability and sex carefully controlled.

Perhaps the better test of the criterion is one which takes place in a "white noise" environment where many (or some) other variables are in operation, rather than in a controlled environment where two variables are purposely highlighted while others are systematically eliminated.

The research (e.g. Smedslund, 1961; Flavell, 1963) strongly suggests that training is highly unlikely to accelerate Piagetian stages of development unless the student is transitional and in a state of disequilibrium. From this, one might hypothesize that developmental gains will occur from instruction in terms of the percentages of students who are transitional prior to instruction.

Unless such experimental research is conducted, we will have to wait for massive amounts of correlational data for a true picture of the relation

between curricular practices and progression through developmental stages.

REFERENCES

Ashley, J. and D. Butts. "A Study of the Impact of an In-service Education Program on Teacher Behavior." In D. Butts (Ed.) Research and Curriculum Development in Science Education, Science Education Center, University of Texas, 1971.

Crowther, F. "Factors Affecting the Rate of Adoption of the 1971 Alberta Social Studies Curriculum for Elementary Schools." Master's thesis, University of Alberta, 1972.

Flavell, J. H. The Developmental Psychology of Jean Piaget. Princeton: D. Van Nostrand, 1963.

Fullan, M. and A. Pomfret. "Research on Curriculum and Instruction Implementation." Review of Educational Research, 47:35-397, 1977.

Kerlinger, F. N. Foundations of Behavioral Research. New York: Holt, Rinehart, and Winston, 1964.

Linn, M. C. and H. C. Thier. "The Effect of Experiential Science on Development of Logical Thinking in Children." Journal of Research in Science Teaching, 12:49-62, 1975.

Persall, J. "An Assessment of the Instructional and Organizational Innovations Implemented in a Selected School System." Unpublished doctoral dissertation, Auburn University, 1972.

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

Quinn, Mary Ellen and Kenneth D. George. "Teaching Hypothesis Formation." Science Education, 59(3):289-296, 1975.

Descriptors--*Educational Research; *Elementary Education; *Elementary School Science; Elementary School Students; *Filmstrips; Grade 6; *Instruction; Science Education

Expanded abstract and analysis prepared especially for I.S.E. by David R. Stevenson, Truro, Nova Scotia.

Purpose

Quinn and George evaluated the effectiveness of teaching hypothesis formation to grade school children. They were interested in the usefulness of the method employed, the measurability of hypothesis quality, and if quality differences were discernible among groups isolated for analysis.

Rationale

The investigation is considered an extension of work by Atkin (1958), and is likewise done in a classroom setting.

Research Design and Procedure

Quinn and George developed an Hypothesis Quality Scale (HQS) from which to judge student hypotheses. The meaning of "precise" was also clarified through a set of statements called "Principles of Precision" (PP); the PP was used to support the HQS. Reliability of the HQS was verified by having three science educators judge 50 hypotheses (0.94 agreement).

The study has a treatment-nontreatment format. A total of 176 sixth grade students from four intact classrooms, two from each of two Catholic schools of differing socioeconomic levels, were subjects. One class from each school was in each of the treatment and nontreatment groups, for a total of 88 subjects in each.

All subjects completed the Otis Test of Mental Maturity prior to treatment. The results served as a covariable in an analysis of covariance that was

part of the Campbell and Stanley Design Twelve procedure. All subjects wrote hypotheses on test films before and after instruction was given to the treatment group. Mean scores for hypotheses produced (scored against the HQS) were subjected to analysis for significance.

Treatment consisted of a series of eighteen 40-minute sessions: film loop showing (two sessions) followed by discussion (one session) in rotation. The loops were produced by the Inquiry Development Program in Physical Science. Each showing was followed by discussion, questioning, reshowing, and hypothesis writing. Hypotheses were scored for subjects according to the HQS. The separate discussion sessions were aimed at improving hypothesis formation.

Before treatment subjects had been divided into groups:

TREATMENT		NONTREATMENT	
(1) Socio I	(2) Socio II	(3) Socio I	(4) Socio II

Subjects in each of groups one to four were further divided for analysis:

a. high and low intelligence,
b. high and low grade point average,
c. high and low reading ability, and
d. sex.

Thus, eight cells were available for each of four analyses of covariance. Interactions of treatment, socioeconomic level and each of the variables above were considered.

Findings

Quinn and George made the following findings:

1. treatment was effective at the .001 level of significance;
2. no significance was attached to socioeconomic status; for the subgroups significant differences were found in the ability to form hypotheses;
3. between levels of intelligence (.02 level);
4. based on grade point average levels (.001 level);

5. between groups by reading ability (.01); and,

6. based on sex (.02).

Interpretations

The investigators concluded from their findings that hypothesis formation can be taught under the conditions in the study; that the method used can be successful, including the use of the particular film loops with sixth grade children; that quality of hypotheses can be measured using the HQS; that equal instruction allows children of differing socioeconomic levels to successfully learn hypothesizing; and, that the ability to hypothesize is correlated with intelligence, grade point average, reading ability and sex of children.

ABSTRACTOR'S ANALYSIS

The study is a straight-forward investigation that fits within the well-used philosophical statement by Bruner about teaching people when they are ready to learn. But there is nothing profound either in the procedure used or in the outcome. Rather, Quinn and George carried out a reasonable teaching assignment and obtained results they expected. What we do not know is whether or not students reached a level of mastery and so have retained the skill, or if the students could generalize to topics other than science. Perhaps, however, such thoughts elevate the importance of the study far beyond that intended by the investigators.

There have been many investigations in the literature of science education about teaching procedures. Indeed it is possible to find support for almost any reasonable position a reviewer may wish to take. However, blinders are necessary if one wishes to claim support for one position only. Quinn and George do not make such a claim; rather, they state simply that the study supports the claim that the method works.

Many studies that are reported in science education journals have nagging minor deficiencies, either as a result of reporting weaknesses or of

features of the studies themselves. The study here under review is no exception. Following are some concerns.

It would be helpful to know the circumstances under which the three science educators undertood their validation of the HQS and who the persons were. Often the captive talents of graduate students are pressed to service and instructions are so carefully given that expected results are obtained. The report level of interjudge agreement suggests no difficulty in this case.

Use of intact classes for research purposes, while a convenience, leaves questions unanswered. How were these particular classes in these particular schools chosen and why? Would different results be obtained from different classes? The study may not have been affected by the use of other classes, but we are not able to judge the range of intelligence, reading ability and achievement and achievement in the sample. Even so, a reviewer would be on the wrong track to overdo criticism of the points in this study, for the outcome is reasonable and expected.

That the film loops were useful and aided in the instruction was shown by the study. Availability of the loops to most sixth grade teachers seems unlikely. In that event, the teacher would be left with the discussion sessions and possibly with much less motivation than that available to the investigators. The reviewer must wonder how much farther we are ahead in a practical sense as a result of the study.

The levels of intelligence, grade point average, reading ability and sex receive no comment other than in the summary of findings. Is it not likely that the differences in hypothesis formation would be found? It seems evident that intelligence, grade point average and reading ability are interrelated and may be measures of similar, if not the same, variables. On the other hand, the difference in sex of subjects opens a question of maturity that could be naturally reflected in the results regardless of treatment. It would be interesting to have the question of differences in mental manipulation by students explored on this point alone.

Quinn and George reported here that equal instruction was given to each socioeconomic group. This is a difficult point to prove. Which group first received instruction? Did that group have an advantage or

disadvantage compared to the other? Does a class of 50 students benefit more or less from "equal" instruction compared to a class of 38 students? At best we may conclude that the two rankings by socioeconomic status were instructed so that no significant difference was found in hypothesis formation.

The sample for this study was not a large one, given the subgroups required for analysis. Should one try to draw conclusions from cells of fewer than 20 subjects? Many reported studies share the disadvantage of small samples or parts thereof subjected to high level analysis. Quinn and George seem in the same position here.

Even though the deficiencies exist, it may be concluded that the results are reasonable and the procedure fair. The outcome is not likely to affect science education to any extent but may assist other investigators who propose to explore instructional procedures in classroom settings.

ATTITUDES AND VALUES

Butzow, John W. and Alan Davis. "The Development of a Semantic Differential of Teachers' Attitudes Toward Teaching Elementary School Science." Science Education, 59(2):211-220, 1975.

Descriptors--Elementary Education; *Elementary School Science; Educational Research; Inservice Teacher Education; *Measurement Instruments; Science Education; *Teacher Attitudes; Teacher Behavior.

Expanded abstract and analysis prepared especially for I.S.E. by Ann C. Howe, Syracuse University.

Purpose

The purpose of this study was to develop and test an instrument for predicting teachers' relative levels of success in using a student-centered, as opposed to a teacher-centered, elementary science curriculum. A semantic differential was developed and used to assess teachers' attitudes toward teaching the Elementary Science Study (ESS) program.

Rationale

This work rests on the assumption that a teacher's level of success in using a particular set of curriculum materials may be predicted by determining whether the teacher has positive attitudes toward the behaviors which the developers of the curriculum consider desirable or necessary. In this case the indicator for level of success was the degree of "student-centeredness" achieved by a teacher.

The reasons for choosing the semantic differential are not explicitly stated but there is ample precedent for using this technique to measure attitude. A number of previous studies are cited.

Research Design and Procedure

A semantic differential consists of a set of concepts, each followed by the same set of polar adjective pairs arranged at opposite ends of a scale. The concepts may be thought of as stimuli to which the subjects respond by

rating each concept on each scale. In this study three initial concepts related to teaching were rated on 46 adjective pairs by 104 college students enrolled in an elementary methods course. From factor analysis of these data four major factors (valuing, enjoying, striving, and difficulty) emerged. Adjective pairs representing these factors were used as scales for the final instrument. The initial concepts were discarded and a new set of concepts chosen by searching the writings of persons who had been active in the development of ESS. The final instrument, called Semantic Differential Test of Teacher Attitudes (SDTTA), contained 21 concepts and four scales.

The ability of the instrument to predict behavior was tested by administering it to 29 elementary teachers who had recently completed an inservice course and were ready to begin implementing ESS. Several months later, after implementation was underway, the teachers were videotaped while teaching science in their classrooms. The videotapes were analyzed using the Science Curriculum Assessment System Teacher (Matthews and Philips, 1968), an instrument which yields scores on a Teacher Directed Index. A low score on this index indicates behaviors congruent with the expectations of the developers of ESS. In order to determine whether the attitudes, as measured by the SDTTA, were correlated with the observed behaviors, as measured by the Teacher Directed Index, the scores on the two instruments were compared. The set of scores above the median on the Teacher Directed Index were labeled Teacher Directed Group and scores below the median were labeled Student Directed Group. Scores for the first three scales (omitting the adjective pairs representing "difficulty") for each concept of the SDTTA were also split at the median. Frequencies of scores thus obtained were, presumably, entered into a 2 x 2 contingency table from which Chi square values were calculated. Six concepts yielded a chi square value significant at the 0.1 level.

Finally, the concepts of the SDTTA which were found to give a significant Chi square value (above) were used to calculate a D-square value for each participant. D-square represents a "measure of dissimilarity of participant from the average scale scores on the Student Directed Group." Participants were rank ordered on the basis of D-square values and on the basis of Teacher Directed Index scores. A Spearman Rank Correlation coefficient of 0.79 was obtained from these data.

Findings

It is somewhat disconcerting to find that the data of greatest interest are not reported. One looks in vain for average scores on the SDTTA items of the two subgroups, the data used to calculate chi square, or the chi square value obtained.

The principal findings were the following:

1. Of the 21 concepts on the SDTTA there were six which discriminated, at the 0.1 level and on at least one scale, between teachers whose behaviors were judged to be student directed and those whose behaviors were judged to be teacher directed. The six concepts and the scales were:

   A. For me, allowing children to mess around with water is... (Powerful/Powerless)

   B. For me, keeping live plants and animals in the classroom for use in experiments is... (Important/Unimportant; Powerful/Powerless)

   C. For me, being able to correctly answer student questions in science is... (Enjoyable/Unenjoyable)

   D. For me, allowing children to work in groups to discuss their different points of view and findings is...(Enjoyable/Unenjoyable)

   E. For me, teaching science is... (Enjoyable/Unenjoyable)

   F. For me, having a strong background in conceptual and factual science is... (Enjoyable/Unenjoyable)

   On the first concept the Student Directed Group gave a less positive rating than the Teacher Directed Group on potency; that is, the Student Directed Group rated the concept as less powerful. On the other five concepts the Student Directed Group gave more positive ratings.

2. The correlation between scores on these six concepts (above) and scores on the Teacher Directed Index, using the Spearman Rank Correlation method, was statistically significant at the .01 level.

Interpretations

The investigators interpret these results as showing that they have developed an instrument which may be used to predict the degree of teacher directed activity or, conversely, the degree of student-centeredness in a given teacher's classroom.

ABSTRACTOR'S ANALYSIS

The semantic differential technique was developed by Osgood, Suci and Tannenbaum (1957) as a method of measuring the meaning which people attach to concepts. Every concept is thought of as having a semantic space which may be explored by means of this technique. One of the dimensions of this semantic space is attitude, defined as an inferred state of an organism which gives it a predisposition to make an evaluative response to a given stimulus. Attitudes have direction and intensity; they are sometimes referred to as tendencies of approach or avoidance. Most social scientists agree that the way a person behaves in a situation depends upon what the situation means or signifies to him or her and that attitude is an important factor in any social activity, of which teaching may be taken as an example. In terms of the semantic differential, attitude toward a concept is defined as the projection of the meaning ascribed to a concept onto the evaluative dimension of the space.

Another way to say this is that a person's behavior depends to a great extent on his or her attitude; if we could measure attitude then we could, to some extent, predict behavior. This is what the investigators have tried to do; they have sought to find a way to measure attitude in order to predict behavior. For this purpose they constructed and tested a semantic differential.

Seen in final form, a semantic differential is a deceptively simple instrument which belies the care with which the concepts and scales must be chosen and tested. The procedure described by Osgood et al. starts with the selection of preliminary concepts and polar adjective pairs. The latter are adjectives which have opposite meanings; for example, good/bad; hot/cold; clean/dirty. The concepts and adjective pairs are submitted to

an appropriate group who rate each concept on each adjective pair, using a scale which customarily goes from 1 to 7. The adjective pairs are then submitted to factor analysis and, if they have been well chosen, three dominant factors become apparent. In most cases these can be defined as evaluation, potency, and activity. Two or three adjective pairs are then chosen to represent each factor and these form the scales of the final instrument. The preliminary concepts are then replaced by a new set of concepts to be rated on the final set of scales. When the semantic differential is to be used to measure attitudes, only the evaluation factor is represented by the adjective pairs since attitude is thought to be most closely related to this factor.

The developers of the instrument presented in this article followed these procedures in many respects but departed from them in other ways for reasons which are not explained. We are not told, for example, why all four factors were retained in selecting representative scales, rather than only the evaluation factor which is the one associated with attitude. And if all four factors were to be retained, it would have been the usual practice to select two or three adjective pairs to represent each factor rather than only one for each factor. A minor point which is puzzling nevertheless is why five rather than seven points were used on the scales. The authors could probably supply reasonable explanations for each of these departures from the original method; often trivial and not-so-trivial questions are left unanswered due to space limitations or other good reasons. A description of the development of a test is more useful to others, however, if such departures from accepted procedure are noted and explained.

More important than any of the above considerations is the question of whether the SDTTA serves the purpose for which it was intended. At this point a digression seems in order to commend the authors for carrying the development of the instrument beyond the initial stage. Too often we are offered "instruments" which have been designed with some care but are presented prematurely and left for others to validate (or ignore).

Several serious questions must be raised about the usefulness of the SDTTA in its present form. The first of these concerns the small number of items for which a correlation could be shown, even at the marginal (0.1) level of

significance, between the predicted behavior and the observed behavior. Since 21 concepts and three scales were used in the analysis, the total number of items, as items are defined in Osgood et al. (1957; p. 80) is 63. Seven of these were found to be significantly correlated with observed behavior but one of these was in the direction opposite to that predicted. That leaves six items, of the 63, which could be used to predict behavior.

Another question might be asked about the reliability of the scores. Do we know that a teacher would give the same ratings to the concepts on a second administration? Or, to put it another way, is there evidence that the attitudes are stable over time?

A third question concerns the meaning and use of the D-square values. This is a murky area which careful reading of the text and reference to Osgood et al. have failed to clarify. It is not clear to the reader who is not an expert how the D-square values were obtained and what they mean. Were all six concepts, including the one which showed an attitude in the "wrong" direction, used in the calculation? Were dissimilarities from the average score measured without regard to direction? Without more explanation it is difficult to evaluate the significance of the correlation coefficient. This may be a question for which the authors have a ready and convincing answer but many readers will need more details or a reference in order to understand the basis and use of this procedure.

As it stands, the SDTTA does not seem to be quite ready for use in making serious predictions of a teacher's level of success in implementing an activity-centered elementary science curriculum since only about one-tenth of the items have been shown to be useful in predicting teacher behaviors. This is a promising beginning for an instrument which might find many uses in both preservice and inservice teacher education, but more needs to be done to make it a reliable and usable tool for research and training.

REFERENCES

Matthews, C. and D. Phillips. Science Curriculum Assessment System, A Handbook. Tallahasse: Dept. of Science Education, Florida State University, 1968.

Osgood, C. E.; G. J. Suci; and P. H. Tannenbaum. The Measurement of Meaning. Urbana, ILL: University of Illinois Press, 1957.

Maddock, M. N. "The Attitude of Papua New Guineans Towards Investigation, Control and Manipulation of Natural Phenomena." The Australian Science Teachers Journal, 21(1):86-92, 1975.

Descriptors--*Attitudes; *Cultural Differences; Developing Nations; Educational Research; Educationally Disadvantaged; *Physical Environment; *Science Education; *Scientific Attitudes

Maddock, M. N. "The Culture Gap--What is Formal Schooling with its Science Education Component Doing to Papua New Guinea Society." The Australian Science Teachers Journal, 21(1):93-97, 1975.

Descriptors--*Cultural Differences; Developing Nations; Educational Research; *Educationally Disadvantaged; *Science Education; Scientific Attitudes; *Social Attitudes; *Stereotypes; Secondary Education

Expanded abstract and analysis prepared especially for I.S.E. by Hans O. Andersen, Indiana University, Bloomington, Indiana and John R. Staver, DePaul University, Chicago, Illinois.

Purpose

The major purposes of these two articles were to:

1. Report on the development in four languages (English, Pidgin, Hiri-Motu and Enga) of an attitude scale named "The Environmental Phenomena Attitude Scale."
2. Determine if scores on the scale can be attributed to the educational level, region of origin, life style (urban or rural) or sex of the study sample.
3. Determine if there is a significant difference between the mean scores obtained by students on the Environmental Phenomena Attitude Scale, and the mean of the score predicted by the student for uneducated people.
4. Determine if there is a significant difference between the mean scores predicted by the students for minimally educated subjects and the mean scores actually obtained by them.

Rationale

Developing countries often face the fact that formal schooling changes the attitude of students extensively, and that these changes have the

potential for creating a major gap between the educated and uneducated which can lead to alienation and/or confrontation. The purpose of this study was to determine if significant attitude differences existed between high school students and villagers with respect to beliefs about man's ability to investigate, manipulate, and control natural phenomena. Furthermore, because the author wished to explore attitudes of a broad spectrum of Papua New Guineans, versions of the instrument were developed in English, Pidgin, Hiri-Motu and Enga (Raiapo dialect) languages. It was assumed that data obtained from these studies would be useful for educational planners because the documentation of extensive attitude differences could suggest a need for extensive adult education.

Research Design and Procedure

The study involved instrument development, validation, and hypothesis testing.

Instrument Development

The "Environmental Phenomena Attitude Scale" was developed in Papua New Guinea, by means of a number of field trials. The instrument consists of 26 items structured as an interview schedule. The items constructed were based on five attitude related objectives stated as aims of the Papua New Guinea, science syllabus. These five objectives were related to investigation of natural phenomena by observation or experiment, the use of experimental results and observations, technologies as man-made developments, the need to test cultural models or theories and the need to change models and theories in light of new discoveries. A person who believed or had faith in the susceptibility of the universe to human ordering and understanding was described as having a positive attitude towards investigation, manipulation, and control of natural phenomena. The establishment of reliability of the instrument, inter-observer reliability of interviews, and validity of the instrument were described in previous efforts by Maddock (1973, 1974). An instrument development section is included here to describe additional evidence supporting the argument for the instrument's construct validity. Construct validity is supported when one can correctly predict test scores from theory or theory-

like statements. In this case anthropologists' conclusions were that the attitude scores of villagers (on instruments such as the Environmental Phenomena Attitude Scale) who were not formally educated in school should be significantly lower than the attitude scores of formally educated high school students.

A highly significant mean score difference (13.6 points) in favor of the educated student provided support for the construct validity argument.

Hypothesis Testing

The investigator reported using 2X2X2 analysis of variance design to test the effects of education (high school student vs. villagers), region (coastal vs. highland), and life style (urban vs. rural). The only significant mean score differences (13.6 points) reported were between levels of education and the difference favoring the educated was significant beyond the 0.001 level with an epsilon (E) of 0.78.

Analysis of variance was used to explore differences between males and females but no significant differences were reported. The author also reported that no significant differences were found between the two major ethnic groups in the sample, the Enga and Chimbu.

In open-ended discussions following the formal questioning, the students made frequent reference to the importance of science in developing the attitudes they held. The students were also asked to predict which answers they thought the uneducated villagers would give and the reasons for their predictions.

Maddock employed a t-test to determine if the difference between the mean of scores obtained by the students on the attitude scale and the mean of scores predicted by students for the minimally educated was significant. A highly significant ($p < .001$) 21.6 point difference favoring the formally educated was found. The Pearson product moment correlation between the students' scores and the predicted scores was calculated to explore the relationship between the students' scores and the scores predicted by the students for the minimally educated. A significant negative correlation

($r = -0.26$) was discovered which led the investigator to conclude that a tendency existed for students scoring high on the scale to predict scores lower on the range than predictions made by students scoring lower. Maddock suggested that this indicated that the more positive the student's score the more likely he/she was to stereotype his village "wantoks" as negative in attitude.

A t-test was also used to determine if the predicted score means for the minimally educated and the mean of scores actually obtained by them differed significantly. The difference was found to be highly significant ($p < .001$), indicating that the students predicted significantly lower scores for the villagers.

In open-ended questioning of students that followed the formal interview information was collected that allowed the investigator to make the following conclusions:

1. Many students claimed that they did not know the practices and beliefs in their area.
2. Many students patronized and stereotyped village thinking.
3. Many students felt that only students were clever enough to deal with modern technology.
4. Many students had become so indoctrinated with the dogma of standard school explanations that they were not willing to try out new ideas.

Interpretations

The investigator concluded that the "Environmental Phenomena Attitude Scale" was a valid reliable instrument and that it was useful for comparative studies. Also, significant attitude differences existed between high school students and villagers. Further, the formal educational system has had a significant degree of success in shifting attitudes toward the positive end of the continuum; and that educational planners rightly should be concerned about the significant attitudinal gap between the student and the minimally educated.

ABSTRACTOR'S ANALYSIS

Developing countries do not have either funds or expertise to plan and develop curriculum uniquely fitted to their specific needs. Even a modest localization of curriculum is in many instances more than can be managed by local school authorities. Their alternatives are two: 1) Continue their illiteracy or 2) Use materials developed in the Western world. Illiteracy must be eliminated. Therefore, students are rushed into curricular studies that are radically different from anything the adult population has experienced. What effects will this action have on the culture? Developing countries around the world are modernizing their schools and their curricula. Very few studies of the effect of modernizing efforts have been made. Therefore, the Maddock effort which is one of the few studies of the local effects of new science curricula in a developing country, becomes quite important.

Professor Maddock's initial effort was the development of the Environmental Phenomena Attitude Scale to the point where it could be reliably administered by interview in four languages. The development of the instrument was only briefly described in these papers because it was the subject of earlier papers (1,2). Persons finding Maddock's effort interesting will want to study these earlier papers. The fact that students demonstrated a significantly more positive attitude towards investigation, control and manipulation was reported because it supported a construct derived from anthropological studies which thereby became additional evidence supporting the construct validity of the instrument.

The fact that significant differences in scores could not be attributed to a regional effect (coastal vs. highland) or an urban-rural effect was at least partially explainable. Maddock indicated that serious attempts were made to make sure that the samples were as representative as possible. However, he reported that it was necessary to use volunteers. It is possible that volunteers would volunteer because they already had a more positive attitude on the dimensions being studied. That being a possibility, it may be safe to infer that the differences in attitude between the students and the minimally educated may be even greater than the highly significant difference reported. Hence, one may well want to encourage

early implementation of adult education efforts under the assumption that gaps between generations are too wide and could lead to the development of social problems.

The first study supported a claim that the educational system (not only science) was influencing the development of a more positive attitude towards investigation, control, and manipulation of natural phenomena than was the lack of formal school education. The second study may well be more significant in that it permitted the conclusion that there was a significant difference between students' scores and scores that they predicted for the minimally educated and a significant difference between the students' predicted and actual achievement scores of the minimally educated. Examination of the means permits the conclusion that the students predicted (incorrectly) that the minimally educated would score significantly lower. When this is coupled with other student interview information which caused the investigator to conclude that the students' descriptions of villagers "revealed a staggering degree of patronizing stereotyping of village thinking," one can understand the cause for concern.

There is evidence to suggest that schools cause students to view at least these dimensions of science more positively than the minimally educated person. Hence, in that respect, the schools can be credited! However, it can also be argued that the schools fail to help the students visualize and understand the positive aspects of their culture. Hence, the schools are failing.

Certainly one must support the preparers of the five-year plan who are calling for adult education. One should also demand revision of student curricula to include intensive study of the local culture, especially those aspects of the culture within the dimensions of this study.

There is a gap, and it was beautifully exposed by the investigator. If eliminating this gap becomes an objective, one would hope that all remember that both the students and the minimally trained must become the objective of training.

Few countries will be able to offer massive curriculum development efforts. Economically, at least superficially, adapting and adopting Western-developed curriculum may be more prudent. However, curriculum developers in developing countries are urged to proceed with caution. Above all, we must remember that the existing culture has been successful. It is the product of local evolutionary processes. It has, until recently, adapted. Parts of the culture, probably large parts, must be preserved.

Finally, we should mention future research directions. This study illustrated one effect of change that was only partially beneficial and exposed dimensions that need treatment. Similar studies should be conducted wherever innovations are replacing existing cultural practices.

REFERENCES

Maddock, M. N. "A Pilot Study to Develop, Refine, and Field Test an Instrument to Measure the Attitude of Papua New Guineans Towards the Investigation, Control and Manipulation of Natural Phenomena." Unpublished Ph.D. dissertation, Florida State University, 1973.

Maddock, M. N. "An Attitude Scale for Papua New Guina." The Australian Science Teacher Journal, 20(2):83-89, 1974.

Huston, Peter H. "A Study of Value Orientations as a Characteristic of Secondary School Students and Teachers of Chemistry." Journal of Research in Science Teaching, 12(1):25-30, 1975.

Descriptors--*Chemistry; Humanism; *Relevance (Education); Science Education; Secondary Education; *Secondary School Science; *Teacher Characteristics; Technology; *Values

Expanded abstract and analysis prepared especially for ISE by Elizabeth Kean, University of Wisconsin.

Purpose

The purpose of this study was to investigate the value orientations (preference for humanistic, technological or theoretical aspects of science) of teachers and students of chemistry with respect to characteristics of those groups. The following relationships were explored:

1. the correlation between value orientations measured on a developed instrument and student characteristics of measured intelligence, general academic average, and chemistry grade;
2. differences between the means of the value orientations for male and female students;
3. differences between the means of the value orientations for teachers and students;
4. differences in the mean value orientations of teachers related to levels of university chemistry preparation and levels of teaching experience;
5. differences in the mean value orientations of those teachers with a maximum of five university chemistry courses who were teaching biology at least quarter time and those teachers with similar preparation who were not teaching biology at least quarter time.

Rationale

This work does not appear to be specifically related to other previous research. It rests on the assumptions that science (i.e., chemistry) curricula should reflect the value orientations of students, and that the science (chemistry) curriculum is heavily based on a theoretical orientation.

A further assumption is that it is possible to separate and measure the humanistic, technological, and theoretical aspects of science.

Research Design and Procedures

The author used a survey design (non-experimental and non-quasi experimental) in an attempt to describe a part of the educational context.

Instrument

The author designed, tested, and utilized a forced choice ranking instrument to measure value orientations of teachers/students. The instrument consisted of an unspecified number of statements of chemical fact, each of which was followed by three further related (correct) facts. Of the three related facts, one had a theoretical, one a humanistic, and one a technological emphasis. From the three examples listed in the article, the orientation statements appeared to be placed in random order.

First preference rankings were accorded a score of two points, and second preference a score of one point. Students/teachers were instructed to rank each set of alternatives according to their personal preference or idea of importance, assuming all statements were correct. Points were summed for all sets of alternatives, providing three scores for each person which indicated the strength of their orientation to each value.

Test items were pretested on a "similar" (but unspecified) population. Content validity was checked by a panel of five experts (unidentified) at an 80 percent agreement level. Construct validity was supported by mean value orientations reported for groups of theology and engineering students (theology students scored higher on humanistic and lower on theoretical orientations; scores on technological orientations for the two groups were similar). Moderately high statistical reliability was reported for teachers and for students and each orientation.

Sample

The student sample consisted of 60 girls and 60 boys in a suburban London, Ontario, school. All students were from the middle socioeconomic level, had

had three years of high school science, were in the upper two-thirds of their classes academically, and were about the same age (i.e., all were 12th grade). Data on students (from school records) included IQ scores, grades in chemistry, but not previous science grades.

The teacher sample consisted of all 39 teachers teaching chemistry in the above school system in the 1970-71 school year. Teachers provided self-report data on years of teaching experience, number of university level chemistry courses, chemistry and biology teaching assignments. No breakdown on length of time since university work, age, or sex of teachers was reported, nor was there any indication of teacher participation in subject matter assignments.

Procedure

The instrument was administered to students during one class period at an unspecified time of the school year. No information was provided regarding the procedure for administration of the instrument to teachers. No description of the chemistry curriculum in use in the system was provided.

Correlation coefficients between value orientation scores and student characteristics were calculated. Differences in the means for various blocking variables were tested for significance with a t-test.

Findings

No raw data (tables of correlation coefficients, means for groups, standard deviations) or inferential statistics were included in the paper. The author did report the following qualitative findings:

1. "No correlation" between value orientation and measured intelligence, general academic average, or chemistry grade for students.
2. "A significantly higher" orientation of girls to humanistic and boys to technological orientations.
3. Students were "more highly" oriented to humanistic and technological values than were teachers, who preferred the theoretical.

4. When teachers were blocked on two levels of experience and two levels of chemistry course work:

   --higher university preparation was associated with higher theoretical and lower humanistic and technological orientations, and vice versa.

   --higher level of teaching experience was associated with lower theoretical and higher humanistic and technological orientations, and vice versa.

   These results were not statistically significant unless both factors were combined.

5. Teachers who taught at least one quarter time biology had lower theoretical and higher humanistic and technological orientations than had the teachers who taught only chemistry.

Interpretations

The following interpretations were reported:

1. Value orientations of students must be measured directly since they do not correlate with student characteristics.
2. Sex differences for students conform to a common stereotype; the source is still undetermined (societal expectation vs. fundamental difference).
3. Chemistry curricula coincide more closely to value orientation of teachers rather than with those of students, implying a need for revision to match those of students.
4. No causality could be inferred from results of teaching experience and level of preparation since these factors might be confounded.
5. No implications could be drawn from differences in teaching assignments, since comparisons with other subject matter assignments were not available.

Validity

The author has designed an instrument that apparently can measure the humanistic, technological, and theoretical value orientations of individuals toward chemistry. He made appropriate efforts to control variables in his samples. Some additional information on the samples/instruments might have been useful (see section 3 above), but for the stated purposes of the paper, information was generally satisfactory.

A major validity question arises, however, from the lack of data reported within the paper. Since the author does not include correlation coefficients, means, and standard deviations for groups, levels of significance, etc., the reader is left with questions as to what was actually found and cannot judge whether reported conclusions are actually warranted.

The validity of one assumption is open to question: whether or not curricula must match the value orientations of students. Can it not also be argued that the purpose of education is to intervene in the values of students and to offer them new options? It is not clear what the optimum match in orientation might be in order to develop maximum interest on the part of students. The actual curriculum that students in this study had been exposed to has not been discussed by the author, nor is there any indication what its orientation might be.

Research Design

This study comes under the rubric of descriptive research; i.e., it does not attempt to manipulate an instructional situation but to discover some important and meaningful facets of that situation. As such, it is extremely susceptible to biases inherent in the research questions asked.

This study purports to ask whether students (assumed to be naive learners of science) possess a theoretical, humanistic, or technological orientation with respect to chemical content. Their responses to the forced choice rankings are then compared to responses of teachers of chemistry. Means for groups

of teachers with different characteristics are also compared with one another. The critical question which comes to mind is: to what extent are responses dependent upon chemical knowledge? To what extent can naive learners of science with little theoretical knowledge and few experiences with chemical theory prefer a theoretical value? The answer seems predetermined: students and teachers with less knowledge and exposure to scientific theory will prefer orientations other than the theoretical when compared to those teachers who have had extensive exposure to the "correct" and "expected" theoretical orientation in college chemistry classes.

The same problem applies to the attempted construct validity in Table 1. Again, exposure to more of the accepted theoretical orientation in science classes by engineering science students might account for the differences in means for them compared to theology students.

There is also a question whether one should expect any correlation between intelligence, general academic average and chemistry grade with a theoretical orientation. What is lacking is a theoretical construct which could explain such correlations, even if found. Further, was there any question in the reader's mind that an abstract, theoretical chemistry curriculum would not reflect teachers' values rather than students? Examples of learner oriented science courses are rare indeed.

In the case of male vs. female student differences in values, the latter's preference for humanistic orientation is readily explained by societal expectations that females will be people oriented. Again, we have learned nothing new. What might have been interesting to find out is whether or not teachers of both sexes with similar academic chemical background differ in their value orientations.

Since no construct is presented that would explain any relationship of teaching assignment to value orientation, this hypothesis contributed little to the paper, and probably could have been omitted.

Adequacy of the Written Report

Some of the difficulties mentioned above may be due in part to the origin of the study--the author's dissertation. What is presented here is a part of a larger work, whose nature is suggested by the thesis title. It may have been pertinent in the thesis to block on teaching assignments, particularly when considering learning outcomes, but the inclusion of this aspect in this shorter work seems questionable. It is difficult to abstract a short paper from a larger and more comprehensive work. The danger is that one abstracts parts without providing sufficient rationale or relevancy to the shorter work.

Suggestions For Future Work

The author's suggestion that alternate forms of the Chemistry Preference Evaluation Instrument might be used to measure changes in value orientation points to potential further research. However, any future work in value orientation of teachers and students should have a clear notion of purpose and value. The results should have some implication for how teaching/learning might be organized or conducted.

Any further work in value orientation and its potential effect on curriculum should have a stronger conceptual base. What questions are worth seeking answers to? Is one orientation better than another? Can value orientations be changed by exposure to more science information? By exposure to different types of curricula? If cognitive level (i.e., knowledge of chemical theory of a theoretical nature) is controlled, how do we explain value orientation preferences? Are these preferences stable over time? Asking stronger questions might generate more educationally significant results.

AUTO-TUTORIAL INSTRUCTION

Nordland, Floyd H.; Jane B. Kahle; Stephen Randak; and Thomas Watts. "An Analysis of the Effectiveness of Audio-Tutorial Instruction: Measured by Student Achievement and Predicted by Standardized Measures." School Science and Mathematics, 277-284, 1978.

Descriptors--*Audiovisual Instruction; *Biology; Individualized Instruction; *Instruction; Learning Disabilities; *Research; Standardized Tests; Secondary Education; *Secondary School Science

Expanded abstract and analysis prepared especially for I.S.E. by Glenn H. Crumb, University of Western Kentucky.

Purpose

The purpose of this investigation was to compare achievement by 119 high school biology students after receiving instruction by either an individualized, audio-tutorial system or by group classroom presentations. It was hypothesized that (a) there would be no difference in achievement on teacher constructed unit tests; (b) there would be no significant difference in the correlation between scores on selected predictor variables and achievement in the two treatment groups; and (c) there would be no significant difference in achievement between the treatment groups as subdivided at the fortieth percentile on national norms for selected predictor variables.

Rationale

The concern of the investigation focuses upon a common problem in most high school science courses—how to effectively deal with student differences in reading ability, verbal ability, and the ability to use quantitative information. Small group or individualized instruction as described by Postlethwait, Novak and Murray (1969) and others, seems to offer at least partial answers to this classroom dilemma, since it has already been established that such instructional strategies can be effective when utilized with a selected population of learners at the college and secondary school levels. When standardized measures employed (Otis I.Q., SCAT, SAT) failed to predict differences in student achievement between the two treatment groups it was hypothesized by the authors that the audio-tutorial method of instruction might be providing an alternative for students deficient in certain skills (reading, quantative abilities).

Research Design and Procedure

Two randomly selected treatment groups (n = 59) were taught mitosis, meioses, probability and genetics using materials developed by the high school teachers in collaboration with the senior authors. During a treatment period of three weeks duration, one group (AT) received all instruction material by an individualized, audio-tutorial format, while the other group (NAT) was taught using a group, classroom format. The hypothesis tested were as follows:

1. (Ho:$\bar{X}AT = \bar{X}NAT$, H_1 :$\bar{X}AT \neq \bar{X}NAT$)
   No significant difference in achievement.

2. (Ho:r AT = r NAT, H_1: r AT ≠ r NAT)
   No significant difference in correlation between scores or selected predictor variables and achievement.

3. (Ho:$\bar{X}$AT 40th% = $\bar{X}$NAT 40th%
   $\bar{X}$AT 40th% = $\bar{X}$NAT 40th%
   H_1:$\bar{X}$AT 40th% $\bar{X}$NAT 40th%: $\bar{X}$AT 40th% $\bar{X}$NAT 40th%)
   No significant difference in achievement between treatment groups as subdivided at the 40th percentile on national norms for each predictor variable (SCAT quantitative, verbal, total; Otis IQ; STEP reading; SAT Language, Arith. app., para. meaning.)

Findings

Hypothesis One: Ho was not rejected as no significant difference in achievement was found using a pooled variance estimate t-test.

Hypothesis Two: Ho was rejected for all predictor variables except the SCAT quantitative as a result of the application of the Fisher z statistic.

Hypothesis Three: Ho was not rejected for every predictor variable in the population above the 40th percentile. Ho was rejected for every predictor variable except the SCAT quantitative in the population at or below the 40th percentile.

Based upon the data and the results of statistical analysis, the following were reported:

1. The audio-tutorial group had higher mean scores on the teacher-made unit tests although the differences in achievement between the two groups so measured was not significant.

2. Standardized measures were less predictive of achievement for the groups taught by the audio-tutorial format.

3. When the treatment groups were subdivided at the 40th percentile level for each predictor variable, no significant differences were found in the ability of the variables to predict achievement for subjects scoring above the 40th percentile.

   There was a significant difference between the treatment groups in the ability of the standardized measures (excepting for the SCAT quantitative) to predict achievement for subjects scoring at or below the 40th percentile.

Interpretation

It was reported by the investigators that the marked contrast in the ability of the standardized measures to predict achievement among individuals below the 40th percentile "would seem to support the individualized audio-tutorial method as a meaningful alternative for students possessing certain deficiencies."

ABSTRACTOR'S ANALYSIS

The preceding statement points up one of the weaknesses of the study—namely, the lack of a discussion of implications of the study in the context of previous work by others. For example, Pinsky (1973) reported that students whose learning activities were controlled the most showed the greatest gains in achievement. Inasmuch as the audio-tutorial procedures may have

considerable control on the learning activities of the AT group members, some discussion of this factor would seem appropriate, particularly in view of the performance of the group at or below the 40th percentile. A further clarification would also seem to be needed regarding the instructional styles and roles of the two teachers in each of the AT and NAT groups.

Perhaps because of space limitations, the authors provide no insight into the instructional strategies employed with the NAT group. It would seem very unsafe to make predictions concerning appropriateness of use of audio-tutorial instruction in lieu of other teaching strategies without clearly delineating the similarities and differences that prevail between the two programs. It does not appear that the two authors are reporting anything new regarding the ability of individualized, audio-tutorial instruction to successfully teach science concepts. It may be, however, that the materials utilized are of a unique design which may provide benefits to some students receiving instruction in some specific areas of high school biology.

The concise manner in which the study was reported obviously leaves some questions unanswered regarding the data analysis. For example, the authors used a t-test "to verify the effectiveness of the randomization procedures." If, indeed, the procedures used for assignment to groups was truly random, no verification would be required! Although this extra precaution was taken, no statement is provided as to which standard measures were used as a base for comparison. If they were the same ones as used for selection of the groups falling below and above the 40th percentile, a substantial burden has been placed upon the validity and reliability of scores on those standard measures.

If this is the case, then questions regarding date of testing, equivalence test forms, the previous student experiences in school (grade level) need to be focused upon. For example, the authors make reference to students in grades 9 through 12. What assumptions were made regarding equality of subject naivete among the AT and NAT group members? How does the analysis account for different experiences of twelfth grade students as compared to ninth graders? Obviously the latter will have had fewer opportunities for instruction in the content areas which were included in the materials used in the study. In short, for a three-week period of training in an area

that is specifically science oriented, predictors of the type used (Otis IQ, SAT, STEP, SCAT) may not be as appropriate when used alone, as would previous experience and achievement in related subject matter content. The final results of the study may be as much due to a qualitative difference in the students as in the AT versus NAT difference. Any trend established prior to the experimental period may have had considerable influence on the results reported.

The criticisms cited in no way are designed to reflect negatively upon the authors or the reported study. Given the constraints of the real world of schools, classes, and policies of research journals, the study as reported does provide some added insight into a substantial problem in the teaching of biology in our nation's schools.

Rowsey, Robert E. and William H. Mason. "Immediate Achievement and Retention in Audio-Tutorial Versus Conventional Lecture-Laboratory Instruction." Journal of Research in Science Teaching, 12(4):393-397, 1975.

Descriptors--*Autoinstructional Methods; *Biology; *College Science; Educational Research; Higher Education; Instruction; Science Education; *Teaching Methods

Expanded abstract and analysis prepared especially for I.S.E. by David L. Dunlop, University of Pennsylvania at Johnstown.

Purpose

The purpose of this study was to analyze the outcomes of two methods of instruction, conventional lecture-laboratory and audio-tutorial, as related to immediate achievement and to retention among students in a university course in animal biology. The following two research questions were examined: 1) Will students in an audio-tutorial program show greater immediate achievement than those in conventional lecture-laboratory programs? 2) Will students in an audio-tutorial program show greater differences in the retention level than those in a conventional lecture-laboratory program?

Rationale

As the authors correctly indicate, the audio-tutorial approach to instruction has grown into a widely accepted teaching method and has been used at every level of instruction--elementary school through graduate and professional school. It appears that the rationale for conducting this study was to provide instructors with some data useful in selecting between two separate formats of instruction.

Research Design and Procedure

The experimental group (n = 134) participated in an audio-tutorial program which consisted of 18 separate exercises which were developed and recorded on cassette tapes. The entire taped lecture sequence required approximately 15 hours, with individual exercises ranging in length from 15 to 85

minutes. The program also included voluntary group discussions once a week and assistance from graduate students or faculty members when needed. The control group (n = 190) consisted of two conventional lecture-laboratory sections, each of which met for four 50-minute lecture sessions and one three-hour laboratory session a week. Students enrolled in one or the other of these two groups without realizing that two separate methods of instruction were available. The afternoon session was randomly selected as the experimental section.

An achievement test consisting of 60 multiple choice items was developed and used as a pretest, posttest, and retention data gathering instrument. A reliability of .86 was determined through the use of the split-half method and the Spearman-Brown formula, and the content validity was established by a panel of qualified individuals. The professors involved in the study did not participate in the construction or validation of the achievement test and they had no knowledge of its contents.

The pretest was administered to all of the students on the first day of the quarter and again at the completion of the quarter. Eleven weeks and three days later the same test was again administered to all of the available students. (Twenty-nine percent of the control group and 34 percent of the experimental group participated in the retention study.) Five t-tests were conducted in the analysis of the data.

Findings

The pretest scores for the two groups were not significantly different; however, the posttest scores and the retention scores were significantly different with the experimental group being higher in both instances.

Interpretations

No specific conclusions were stated. However, it was implied that the audio-tutorial approach to teaching college biology was superior to the conventional methods of instruction with respect to both immediate achievement and retention.

ABSTRACTOR'S ANALYSIS

Today there are several different degrees of audio-tutorial sophistication; however, Postlethwait (1970) states that the audio-tutorial system was begun in 1961 as an attempt to assist Purdue University students who had a poor background in introductory botany. This method of instruction began as a taped lecture and was later supplemented with a textbook, laboratory manual, specimens, experimental equipment and other such items. Postlethwait's system then evolved into a much more elaborate and integrated program, which now includes several study sessions a week, an independent study session, a general assembly session, and an integrated quiz session (1969).

The audio-tutorial program which Rowsey and Mason used as a basis of their study appears to be similar to Postlethwait's earlier program as it does not contain the numerous elements of Postlethwait's later, more integrated approach. Recognizing this difference, one could speculate that a replication of this study using a more integrated audio-tutorial approach would result in even greater differences between the control and experimental groups.

In their review of literature Rowsey and Mason mentioned eight related studies; however, two of these studies (Weaver and Russell) were omitted from the reference list at the end of the article. The absence of a correct reference to Russell's work is especially frustrating as it is the only study mentioned which demonstrated that students taught by the conventional lecture-laboratory method of instruction achieved at a level significantly greater than students instructed by an audio-tutorial program.

As Rowsey and Mason's literature review demonstrated, the audio-tutorial instructional approach has produced conflicting findings, and an expansion of the literature review will not totally resolve the issue. Studies by Meleca (1970, Hamlet and Holt (1973), Hahn (1971), and Bish, Bowman and Sarachek (1978) support the superiority of audio-tutorial approaches over a more conventional method. However, studies by Durst (1968), and by Grobe and Sturges (1973) do not support the superiority of the AT method.

It appears that future investigations comparing the audio-tutorial instructional method with other methods would do well to devote some additional thought to the reasons for the conflicting findings reported in the literature. One approach would be to further refine the research design and apply statistical techniques useful in identifying the variables relevant to a student's learning in these approaches to instruction. Bish, Bowman, and Sarachek (1978) have begun this, and they report that of 18 variables studied only the precourse test score, course section (AT or LL), and class contact time are significant predictors of student performance in the course. A related approach which could also be attempted is that of using Hunt's matching models concept (1971). He describes in detail the necessity of matching an individual with the "correct" educational environment. If this were properly done, we would expect to observe greater achievement in all types of learning programs.

Pare and Butzow (1975) examined the relationships among independence of work habits, attitude, and achievement in an audio-tutorial physical science course. They conclude that audio-tutorial instruction does not appear to be a panacea for all students and that more research on this instructional method is necessary before we will understand all of the related parameters.

Rowsey and Mason seem to ignore variables such as learning styles, attitudes, work habits and etc. when comparing the experimental and control groups. Based upon similar pretest measures (cognitive scores), they state that their two groups were "homogeneous" at the beginning of the study. Were the groups homogeneous with respect to other relevant variables? If not, what effect, if any, would this have on the study?

Since long-term retention is a significant part of one's education, the measurement of retention is indeed important. However, the number of students participating in the retention portion of this study is of some concern.

Why was this number so small? Was there some type of bias operating? For example, could it be that several of the original participants did not return to school in the fall due to low academic records? If so, this could introduce bias and make interpretation and generalization very difficult, if not impossible.

In addition to the suggestions made earlier for future research, one additional area should be mentioned. That is, research related to video-autotutorial instruction. Fisher, Buenther, and Macwhinney (1977) have begun to examine this area and they conclude that video-autotutorial instruction seems to be a particularly effective method of instruction. In view of the increasing availability of VTR equipment, this area of research deserves additional study.

REFERENCES

Bish, John T.; B. L. Bowman; and A. Sarachek. "Lecture-Laboratory vs. Structured Audio-Tutorial Approaches: Student Achievement." Journal of College Science Teaching, 7:168-171, 1978.

Durst, Harold E. "An Evaluation of an Experimental Biology Course Employing Some Cognitive and Noncognitive Achievements as Criteria." Dissertation Abstracts, 28:2438, February, 1968.

Fisher, K. M.; H. Guenther; and B. Macwhinney. "Does Video-Autotutorial Instruction Improve College Student Achievement?" Journal of Research in Science Teaching, 14:481-498, 1977.

Grobe, Cary H. and Allan W. Sturges. "The Audio-tutorial and Conventional Methods of College-level Biology for Nonscience Majors." Science Education, 57:65-70, 1973.

Hacket, Dorothy and Imy V. Holt. "Biological Science as an Audio-Tutorial System of Instruction for the Nonscience Major." Science Education, 57:499-516, 1973.

Hahn, T. C. "Audio-tutorial Instruction: A Case Study.: BioScience, 21:814-819, 1971.

Hunt, David E. Matching Models in Education. Ontario, Canada: Ontario Institute for Studies in Education, 1971.

Meleca, C. Benjamin. "Multiple Linear Regression Analysis: Results and Discussion II.". BioScience, 20:26-30, 1970.

Pare, Roland R. and John W. Butzow. "The Relationships Among Independence of Work Habits, Attitude, and Achievement in An Audio-Tutorial Physical Science Course." Journal of Research in Science Teaching, 12:1-3, 1975.

Postlethwait, S. N. "The Audio-Tutorial System." The American Biology Teacher, 32:31-33, 1970.

Postlethwait, S. N.; J. Novak; and H. T. Murray. The Audio-Tutorial Approach to Learning. Burgess Publishing Co., Minneapolis, 1969.