DOCUMENT RESUME

ED 290 621 SE 048 852

AUTHOR Blosser, Patricia E., Ed.; Helgeson, Stanley L., Ed.
TITLE Investigations in Science Education. Volume 13, Number 4, 1987.
INSTITUTION Ohio State Univ., Columbus. Center for Science and Mathematics Education.
PUB DATE 87
NOTE 65p.
AVAILABLE FROM Information Reference Center (SMEAC/IRC), The Ohio State University, 1200 Chambers Rd., 3rd Floor, Columbus, OH 43212 (U.S. subscription $8.00; $2.75 single copy).
PUB TYPE Collected Works - Serials (022) -- Reports - Research/Technical (143) -- Information Analyses (070)
JOURNAL CIT Investigations in Science Education; v13 n4 1987

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Chemistry; College Science; *Elementary School Science; Elementary Secondary Education; *Environmental Education; National Surveys; Physics; *Problem Solving; Process Education; *Science Curriculum; Science Education; *Science Instruction; Science Laboratories; *Secondary School Science; Student Attitudes; Supervisors
IDENTIFIERS Science Education Research

ABSTRACT

Abstracts and abstractor' critiques of nine published research reports related to problem solving, elementary school science, science process skills, and curriculum and instruction are presented. Aspects addressed in the studies include: (1) problem-solving skills of high school chemistry students; (2) spatial puzzles and the assessment of children's problem-solving performance; (3) high school students' ability to solve molarity problems and their analog counterparts; (4) the development of a children's science measure; (5) teacher and supervisory perceptions of elementary science supervision; (6) student task involvement and achievement in process-oriented science activities; (7) attainment of skills in using science processes: instrumentation, methodology, and analysis; (8) secondary school students' beliefs about the physic laboratory; and (9) a national survey of curriculum needs as perceived by professional envir ..mental educators. Included for each study are the purpose, rationale, research design and procedure, findings, and interpretations. (TW)

INVESTIGATIONS IN SCIENCE EDUCATION

National Association for Research in Science Teaching

ERIC Clearinghouse for Science, Mathematics, and Environmental Education

INVESTIGATIONS IN SCIENCE EDUCATION

Volume 13, Number 4, 1987

Published Quarterly by

SMEAC Information Reference Center
The Ohio State University
1200 Chambers Road, 3rd Floor
Columbus, Ohio 43212

Subscription Price: $8.00 per year. Single Copy Price: $2.75.
Add $1.00 for Canadian mailings and $3.00 for foreign mailings.

Volume 13, Number 4, 1987

NOTES FROM THE EDITOR:

This fourth issue of Volume 13 contains critiques of published research articles on problem solving (three articles), elementary school science (two articles), science process skills (two articles), and curriculum and instruction (two articles).

The Problem Solving section contains critiques of research conducted by Gabel, Sherwood and Enochs; by Hill and Redden; and by Gabel and Samuel. The two articles of which Gabel is the first author relate to research conducted in high school chemistry classes. The research by Hill and Redden involved fifth grade students.

The two critiques of Elementary School Science research involve articles by Harty and Beall who were interested in measuring the scientific curiosity of elementary school children and by Perrine who wished to investigate perceptions of elementary teachers and supervisors of ideal supervisory practices.

The Science Process Skill research involved early adolescents and their engagement time in process learning (Tobin) and early adolescents working with microcomputers to develop skill in estimation (Berger).

The Curriculum and Instruction section contains critiques of two diverse articles: a report of a sucessful effort to increase high school physics enrollment (Renner, Abraham and Birnie) and a report on the assessment of curriculum needs in environmental education at all school levels in the United States (Volk, Hungerford, and Tomera).

Patricia E. Blosser
Editor

Stanley L. Helgeson
Associate Editor

PROBLEM SOLVING

Gabel, Dorothy L., Robert D. Sherwood, and Larry Enochs. "Problem-Solving Skills of High School Chemistry Students." Journal of Research in Science Teaching, 21 (2): 221-233, 1984.

Descriptors--Chemistry; Decision Making Skills; High School Students; Information Utilization; *Learning Strategies; Logical Thinking; *Problem Solving; Secondary School Science; Science Education; *Science Education Research; Transfer of Training

Expanded abstract and analysis prepared especially for I.S.E. by Patricia H. Suter, Del Mar College, Corpus Christi.

Purpose

The primary objective of this study, which was supported by the National Science Foundation, was to compare processes used by students who were successful problem solvers with those who were unsuccessful.

Rationale

The rationale for this study is the prime importance of problem solving to science teaching. What strategies do high school students use when solving chemistry problems? Teachers need to realize the methods students use to successfully solve problems to be better able to transfer these methods to unsuccessful students.

Research Design and Procedure

The sample consisted of 266 high school students who were selected from 609 chemistry students enrolled in the classes of 10 teachers in 8 schools in Central and South Central Indiana in 1979-80. Schools were located in small town/rural areas, moderate size cities and suburban and inner city areas. Students were selected on the basis of their willingness to participate, their availability during times interviews were conducted, the strategy studied to solve problems, and their scores on a proportional reasoning test and unit achievement test.

A proportional reasoning skill aptitude test was administered at the beginning of the school year. Students in each classroom were randomly assigned to one of four strategies for learning to solve chemistry problems. These were the factor label method, the use of analogies, the use of diagrams, and proportionality. Booklets were prepared for teaching chemistry problems by these four strategies for the five topics over which students were interviewed. These topics were the mole concept, gas laws, stoichiometry, and molarity. Students were interviewed between two and four weeks of their completing the unit test for the particular type.

The interviewers were two male doctoral students in science education. Prior to their first interview, a detailed protocol booklet for conducting the interviews was devised and studied. Each interview lasted from 30 to 45 minutes and took place outside the chemistry classroom. Students were provided with a piece of paper, pencil, and a periodic table. The interview consisted of three major portions: a think-aloud warm-up period, a question section, and the problems section. The problems section consisted of presenting three problems of increasing difficulty with the last problem being a transfer problem requiring some original thinking.

A proportional reasoning test developed by Staver was expanded to include four tasks and 21 questions. Its alpha reliability was 0.85. Approximately half of the students interviewed were considered low in proportional reasoning, the other half high.

Scores on unit tests were used to rate students' success in problem solving. Each test consisted of 10 items and contained two or three transfer items.

Tapes of the interviews were coded using a protocol adapted from one used by Nurrenbern. All of the tapes were coded three times, twice by the same rater at two or three week intervals, and then by the principal investigator.

Data were analyzed using three different statistical methods. In cases where only frequencies were available, chi-square analyses were used. Kruskal-Wallis one-way analysis of variance, Mann-Whitney U tests, and data summation were all used in analysis.

Findings

Students who were classified according to different degrees of success on their unit test results used different strategies in solving chemistry problems. For the moles unit, students who overtly used organizing skills and in particular used mneumonics were more successful in solving problems. These students generally used systematic procedures, particularly on moles and stoichiometry problems. Less successful students were less systematic in their approach to solving problems.

Students with high proportional reasoning ability used systematic approaches, but also used algorithmic reasoning strategies more frequently than low proportional reasoning students. These were also the more successful students in problem solving.

An unexpected finding was that a large number of students depended only on algorithmic procedures and gave no evidence of reasoning out the problem.

Interpretations

This study shows that when chemistry students solve problems of varying difficulty in a variety of topics (moles, stoichiometry, gas laws, and molarity) few students used recall techniques, and the majority of students relied on strictly algorithmic techniques rather than using reasoning skills. However, this study also indicates that high proportional reasoning students use reasoning techniques more frequently than do low proportional reasoning students on a large number of problems.

All of the above is interpreted by the authors to indicate that the students really did not understand the concepts involved in the various problems presented them.

ABSTRACTOR'S ANALYSIS

This paper seems to be an interesting addition to the literature on problem solving. The results seem to me to be quite predictable in that the authors found that the students with the higher proportional reasoning ability did better on the tests. However, the findings that the use of algorithmic techniques is pretty universal is somewhat disturbing. They conclude that this is a substitute for understanding concepts.

The conclusions of the study were discouraging in many respects. They do, however, give chemistry teachers a clear message. They need to really concentrate on concepts before trying to teach problem solving. Then, the authors recommend that problem solving be taught in a systematic way.

The importance of laboratory work has been downgraded during recent times. High school chemistry students customarily have little or no experience in cooking, cleaning, or construction which leads to simple facts of chemistry and physics. Their ideas of foodstuffs, medications, detergents and building materials have been corrupted and confused by advertising to the point that special efforts must be expended to compensate for these deficiencies. Practical laboratory work cannot, in my view, be excelled for teaching many real things about chemicals and their behavior.

Once students have some practice in dealing with real chemicals, they can be introduced to the concepts which explain what they can observe. Unfortunately, chemistry is so abstract...strange concepts in a foreign language (mathematics)...that students have great difficulty in understanding what is expected of them. These concepts need as much reinforcement as we can give them for there to be any reasonable expectation that the students will be able to handle problems.

The authors suggest that more time be spent on trying to bring qualitative understanding to the concept before teaching problem solving. Then the steps needed to solve problems should be taught in a systematic way. Most of us try to do this. There will always be those students who cannot handle problems no matter what we do. These students would benefit from qualitative understanding of the concepts which comes from experience in the laboratory.

This study was interesting even though it really showed what one would expect in that the student who reasons in a quantitative way does better on problem solving. It does emphasize the need to teach practical chemistry as a way to teach qualitative understanding of chemical concepts.

Hill, Douglas M. and Michael G. Redden. "Spatial Puzzles and the Assessment of Children's Problem-Solving Performance." School Science and Mathematics, 84 (6): 474-483, 1984.

Descriptors--*Cognitive Style; Educational Research; Elementary Education; *Elementary School Mathematics; *Geometric Concepts; *Mathematics Instruction; Observation; *Problem Solving; Puzzles; *Spatial Ability

Expanded abstract and analysis prepared especially for I.S.E. by Meghan M. Twiest and D. Daryl Adams, University of Georgia.

Purpose

The stated purpose of this investigation was "to examine the information which can be gathered by observing children as they solve spatial problems and to demonstrate that this data is consistent with that obtained from a test designed to measure cognitive style."

Rationale

Problem solving is often a major focus for mathematics and science teachers. Awareness of the current problem-solving performance of pupils, knowledge of problem-solving tasks, and the capacity to match such tasks with pupil ability is required for implementing problem solving activities (Harlen, Darwin and Murphy, 1977). Teachers do not always have access to scores from sophisticated testing procedures that define the problem-solving status of their students. Therefore, it would be beneficial if teachers could gain information about the cognitive style of learners from observing simple classroom exercises.

The Group Embedded Figures Test (GEFT) (Witkin, Oltman, Raskin and Karp, 1971) was used in this investigation to assess cognitive style: working memory capacity, premature closure, and field dependence-independence. It was also assessed through the use of a standard jigsaw puzzle.

Research Design and Procedure

The model used was somewhat clinical in nature and involved two methods of data collection: 1) qualitative, including observations and interviews, and 2) quantitative, including completion times and test scores. A questionnaire was also used and may have been qualitative, quantitative, or both. The dependent variables utilized in the study were scores on the GEFT and completion time in seconds on the jigsaw puzzle. Selected interview, questionnaire, and observational data were utilized in a more qualitative manner. Gender was used as an independent variable to form two comparison groups.

The sample used for this study consisted of 70 fifth grade students attending a public elementary school in a middle class area. The 36 boys and 34 girls had a mean age of 10.5. Subjects individually completed a 35-piece wooden jigsaw which pictured one plant and one animal on a uniformly colored background with no distinguishing features. Pieces were presented scrambled, with the picture side up. Time taken to complete the puzzle was recorded and observations were made as subjects worked. Ten subjects were interviewed afterwards for a follow-up on observations made during the jigsaw completion. A post-task questionnaire about their experience and competency with jigsaw puzzles was also administered.

A Campbell and Stanley (1963) notation of the study is as follows:

O_1 O_2 O_3 O_4 O_5

Where:

O_1 = Score on GEFT

O_2 = Jigsaw completion time

O_3 = Observations during jigsaw completion

O_4 = Jigsaw questionnaire

O_5 = Interview (n = 10)

The results of the GEFT were compared with the performance of subjects completing the "standard jigsaw puzzle" and a correlation coefficient was calculated. Independent t-test procedures were employed to compare the means of both males and females for the GEFT and jigsaw completion times.

Findings

The performance of boys exceeded that of girls on both measures. On the GEFT, th was significant (t = 1.3, p .05) but not in jigsaw completion time (t = 1.3, p .20). Correlation between measures (.50) was significant at the .01 level.

It was also found that:

Pupils with lengthy completion times for the jigsaw puzzle operated more randomly than those with the short completion time.

Better problem solvers used two or three attributes, sometimes simultaneously, while those less efficient used one.

Some subjects seemed to be able to hold an image of the missing piece in their minds, while others could not.

Some pupils were more flexible than others in their approach to the jigsaw completion.

Interpretations

The researchers concluded that the significant correlation obtained between jigsaw performance and scores on a test of cognitive style indicate that the two were related. In a discussion of these conclusions, the investigators suggested that the use of spatial puzzles, such as jigsaws, could provide teachers with a valuable opportunity to assess aspects of the problem-solving performance of pupils in a non-verbal context. Thus, information usually available only from formal tests of cognitive style could be readily obtained by observing pupils as they completed jigsaw puzzles.

ABSTRACTOR'S ANALYSIS

This study seemed to be well researched and addressed an important area of education. The actual proceedings of the study, however, raised several questions concerning definition of terms, descriptive methodology, selection procedures, and statistical analysis.

Several of the key terms in t' study lack sufficient definition and explanation to make them completely understandable. The authors' selection o. the GEFT as a measure of cognitive style seems ' be well supported by the literature relevant to the EFT (Embedded Figures Test), but there is no mention as to whether these two tests are actually the same. If the two are the same, this should be made clear. If they are not, the differences and possible implications should be explained. Another term never fully clarified or explained was a "standard jigsaw puzzle." While a good case was made for using puzzles as spatial tasks, no mention was made as to how to select or define a standard jigsaw puzzle. It would have been helpful for the authors to discuss several puzzles and to give an indication as to what was "standard" about puzzles and why they chose this particular puzzle.

Lack of description in methodology also leaves the reader with unanswered questions. How the GEFT was scored and how to interpret these scores is not indicated. When, and under what conditions, the test was administered is also unclear. We presume it was done before the student completed the jigsaw puzzle (see Design section). A time line is not evident and would clarify the procedures section. Information concerning how age and sex may be related to cognitive development would also strengthen the evidence of the study.

As the GEFT scores were reported in Table 1, there is some confusion as to how the table should be read. It may be that the label "GEFT" is misplaced. There is no clear indication as to what scores are GEFT and what unit of measure was used. Errors such as this one could easily lead to confusion or misinterpretation of the data.

The authors fail to inform the reader about their particular selection procedures for interviewing students. Ten of the seventy students were chosen to be interviewed about their jigsaw puzzle strategies. Which students were chosen and why they were chosen is not discussed. Were these students and their strategies representative of the population, or were they examples of the extremes?

A questionnaire was given to all subjects after they had completed both the GEFT and the jigsaw puzzle. This questionnaire was used to assess the subject's prior experience and competency with jigsaw puzzles. No mention as to the outcome of this questionnaire was given in the report. It would have helped the interpretation of the study if the authors had shared the results of the follow-up questionnaire about the subjects' experience and competency with the jigsaw puzzles. It might have offered an avenue for an interesting comparison between experienced and naive subjects and aided in application to classroom use. It also might have guarded against any bias in different abilities or interests. Its use after the fact, with no mention of the results, is "puzzling."

The authors include a large section on tangrams. While these puzzles have factors in common with jigsaw puzzles, we are not sure why the report includes such a lengthy presentation and discussion. If it had been the desire of the authors to make a specific point about tangrams, data should have been collected using them along with, or instead of, jigsaw puzzles. If data on tangrams had been presented, the expenditure of space would have been justified.

Various questions arise concerning the statistical information presented in the article. The authors reported t-test values as a test for significant differences between mean scores of both GEFT and completion times of the jigsaw puzzles for male and female subjects. We made the assumption that the number of students completing both tasks were the same and wondered why the t-value of 1.3 would be significant (p 0.05) for GEFT scores and not puzzle completion times.

No reliability data were given in the report for either the GEFT or the "standard jigsaw puzzle." There should be published reliability data for the GEFT for a similar group of students. An estimate of the reliability could have been computed using the 70 students who took part in this study. A comparison of the two reliabilities would have been helpful. The authors indicated that jigsaw puzzles had been pilot-tested. This would have been a good opportunity for the reliabilites of several puzzles to be compared. A calculated estimate of the reliability of the puzzle used in the study would have made interpretation of the data a little easier.

The authors reported a significant (p 0.05) positive correlation (0.50) of the scores on the GEFT and the puzzle completion time. It seems to us that the more successful student would have had a higher score on the GEFT and a shorter puzzle completion time. This would indicate a negative correlation if the two were related. The authors' contention that the two are related may or may not be true. A correlation coefficient that is large and significant does indeed indicate the relationship between the variation in the two sets of numbers but it does not indicate that the two are definitely related to one another. There are many instances where two sets of data are highly and significantly correlated but are not related. They are, in fact, commonly related to a third factor.

The problem of matching tasks to children addressed in this study is a worthwhile area of research. The idea of moving away from formal testing and towards teacher assessment of students performing observable tasks is a good one. A more controlled study with the same mix of qualitative and quantitative information would be useful.

REFERENCES

Campbell, D. T. and J. C. Stanley, Experimental and Quasiexperimental Design for Research. Chicago: Rand McNally College Publishing Company, 1963.

Harlen, W., A. Darwin and M. Murphy. Match and Mismatch: Raising Questions: Fitting Learning Experiences in Science to Development for Five to Thirteen Year Olds. Edinburgh: Olive and Boyd, 1977.

Witkin, H. A., P. K. Oltman, E. Raskin and S. A. Karp. Manual for Embedded Figures Test. Palo Alto, CA: Consulting Psychologists Inc., 1971.

Gabel, Dorothy L. and K. V. Samuel. "High School Students' Ability to Solve Molarity Problems and Their Analog Counterparts." Journal of Research in Science Teaching, 23: 165-176, 1986.
Descriptors--*Chemistry; High Schools; *Predictor Variables; *Problem Solving; Science Education; *Secondary School Science

Expanded abstract and analysis prepared especially for I.S.E. by Elizabeth Kean, Chemistry Tutorial Program, University of Wisconsin-Madison.

Purpose

The purpose of this study was to present high school students with solution problems based on the real world (making lemonade) and a corresponding set of molarity problems. Responses to problems would reveal whether students understood the analog tasks and would also identify difficulties with the molarity problems. Five specific questions were asked:

1. What specific difficulties do students have with molarity problems and are these the same as in the analog task?
2. Could using analogs in instruction increase the likelihood of students' understanding corresponding chemistry items? How does student achievement on analog problems compare with corresponding chemistry items?
3. Does success on an analog test predict success on a corresponding chemistry test?
4. Do students retain their ability to solve molarity problems, as measured by an end of year test?
5. Does instruction on molarity concepts increase students' capability to solve corresponding analog-task test items?

Rationale

This work is based on the belief that a major reason why students fail to solve conventional general chemistry problems is that they do not understand the underlying chemical concepts. Three studies were cited to support this belief. The use of analogies to present concepts was proposed as a method of improving instruction, although two of five previous studies cited found teaching of analogies had no significant effect on student achievement. It was assumed in these studies that students understood the analogies.

For this study, the authors assumed that students would understand the analogy used (making various solutions of lemonade). Further, they assumed that comparing students' performance on the multiple choice analog and chemistry tests should reveal whether chemical problem solving was inhibited by misunderstandings of physical phenomena or the inability to apply new terms such as molarity, moles, etc.

Research Design and Procedure

The sample consisted of 497 (out of 619) high school students who completed three tests given over the course of the academic year. Students were enrolled in first year chemistry courses in five schools in Indiana, but no comparisons were made among students in different schools.

Two chemistry tests (Form 1 and Form 2) were written, each containing 12 multiple choice items. The chemistry tests were judged by the participating chemistry teachers to contain typical molarity problems. Corresponding items were then written for the two forms of the test for the analog task (making lemonade). Test items were matched on placement on the page, numbers used, and how foils were determined. Each test item was designated by a label in tables reporting results, but these labels were not explained. One example of a molarity and its corresponding lemonade analog problem was given. The end of year test was constructed by selecting the ten items which were most often answered incorrectly. KR-20 reliability coefficients for the tests were reported (all between 0.70 and 0.82).

Each item on Form 1 of a test had a corresponding item on Form 2 of the test. These questions appeared to differ on a single dimension. For example, Question 1 on the molarity test, Form 1, used mass (grams) as the amount of solute; Question 1 of Form 2 used moles as the measure of quantity of solute. Within a form, questions differed on single or on multiple dimensions. There was no information given on criteria used to categorize problems as "more difficult" or "more complex".

Students within classes were randomly assigned to two groups (Form 1 or Form 2). Analog tests were administered within the first weeks of the school year. Molarity tests were administered within two weeks of completion of the molarity unit (sometime during second semester, time unspecified and presumed to be variable). Students took the same form of the molarity test as they had the analog test. Molarity instruction was not described. The summary test was given at the end of the school year.

The data used for analysis was individual student correct scores on individual questions. Comparisons were made between percentages of students answering questions correctly 1) by group, i.e., between forms, (2) within each form on pairs of related questions, (3) by item on the analog and chemistry test (within forms), and (4) by items on the analog/chemistry tests and the summary test (within forms). Chi-squares were used to determine statistical significance of percentage differences, with significance set at p 0.01.

Findings

The major findings are summarized by questions stated in the Purpose section:

1. What are specific areas of difficulty? Both intercomparisons between forms and intracomparisons within the same form indicated that problems are more difficult if a fraction appeared in the problem and if the problem involved changing the volume of the solution by dilution or evaporation.
2. Could analogs help students understand the chemistry concepts? Results indicated that for straightforward, molarity-type problems, students were more successful in solving analogy problems than corresponding molarity problems. Some 25% of students might improve their chemistry performance on these type of problems if they

were able to see the relationship between the analogy and the chemistry and if they understood the necessary chemical terms. Students apparently did not understand the analog when dealing with more difficult concentration and dilution problems.

3. Can analog task performance predict achievement on molarity tests? Using regression techniques, correlation between analog and molarity tests was significant at the 0.0000 level, and accounted for 18% of the variance.
4. Do students retain their capability to solve molarity problems? Achievement was unchanged on four of seven molarity items on the summary test; achievement on two items increased and one decreased. Thirty-three percent of students retained their ability to solve the "more difficult" problems Twenty-five percent never learned the concept (undefined) at all. About "equal numbers of students" were once able to get items correct and then failed on the other test, and vice-versa, with the authors attributing these results to guessing.
5. Can molarity instruction help analog understanding? It was found that chemistry instruction did not aid many students in understanding analog problems.

Interpretations

The authors state that certain types of chemistry problems are difficult, but the source of difficulties is poorly understood. Chemical terminology, inability to relate chemistry problems to an understood analog, missing basic concepts (e.g., concentrating and diluting), algorithmic problem solving may all contribute to chemical problem solving difficulties. To enhance logical, reasoning problem solving, students should be instructed on prerequisite concepts and shown how these relate to solving chemistry problems.

ABSTRACTOR'S ANALYSIS

There is a certain elegance to the concept of this study. The authors wanted to investigate the relationship of problem solving performance on traditional general chemistry problems to some underlying constructs which have real world analogs. They started with some typical chemistry test questions (Form 1) and for each one constructed a second set of questions (Form 2). The corresponding questions in the Forms apparently differed on a single dimension. Some of the questions within each form also differed on a single dimension, although some differed on multiple dimensions. They then built a parallel set of questions using examples from a real world analogy. By measuring performance on matched questions (between and within forms), one can then infer something about the effect of the dimension on level of difficulty and the presence or absence of some underlying concept. This could lead to a better understanding of the conditions under which use of analogies can lead to more effective learning of required science concepts, which, in turn, could improve instruction leading to better problem solving performance.

In designing their study, the authors attended to many of the details that require attention, e.g., random assignment of students to groups. Given the nature of the data, they used simple, appropriate statistical tests for significance. The details which are not specified, e.g., the specific interval of time between teaching molarity and end of year testing, seem minor. One could have asked for information about how many students were represented in each group, but this too seems unlikely to affect seriously the claims made by the authors.

The paper is presented in a concise, tight, easy-to-follow manner. Several of the tables would have been easier to interpret with changes of labels and a more clear statement of what exactly was being compared (see below). Also, the authors included some conclusions in the findings section, some without supporting evidence. For example, in the findings section, the authors commented on difficulties presented by fractions, evaporation and concentration problems, and stated: "For students who can solve the more difficult chemistry and analog items, these concepts are understood ...", p. 173. This conclusion is plausible, but not supported by the work. Again, these are relatively minor matters, given the overall paper.

One major issue that should be raised, however, relates to the concepts or constructs inherent in each of the problems in each form. There is no clear statement in the article about what information is given in the problem, or what question is being asked of the students. The labels provided in the tables and sometimes discussed in the results give hints as to what the problem solving task is, but little specific information. For example, the problem labeled "Grams" could be a problem in which the concentration and the volume of a solution are given and the amount of solute is to be calculated, in mass units of grams. However, we can't be sure. The Form 2 pair, labeled "Moles," likewise is not described. Lacking such information, we do not know what concepts are actually being compared because the use to which the "gram" or "mole" concept is being put is not specified. The addition of information specifying more precisely the concepts tested in each question would have made Tables I and II easier to comprehend.

A second issue arises in the nature of the concepts within the problems. How do students view these constructs? What do "cups" or "moles" mean to students? Do they see them in terms of a quantity of solute? What do students believe about concentration of solutions, either molarity or taste? Do they see concentration as being independent of sample size, once the solution has been made? Moreover, how good is the analogy? The authors have paired "ounces" with "grams" and "cups" with "moles." Are these concepts truly analogous? Ounces and grams seem reasonably so, although the unit of "fluid ounces," a volume measurement, may cause confusion for some students. Likewise, "cups" of lemonade, assuming this to be powdered concentrate of some form, and "moles" of chemicals are both ways of describing an amount of solute. Are "taste," a qualitative description, and "molarity," a quantitative description of concentration, analagous in students' minds? We are given no information.

Lack of information about the state of prior conceptual knowledge and problem solving performace of the students hinders the ability to interpret the results. Have students ever made a solution of lemonade or a solution in the laboratory? Have students had experiences in diluting or evaporating? If so, what were they? We find out what many students can't do (solve problems involving fractions, diluting and evaporating) but, because of the multiple choice nature of the tests, we have little clue as to why they can't solve these problems. Are students attempting to reason their way through the problems? Do they use proportions or the factor label method? How do the methods compare for the analog and the molarity task? Given a correlation between the tasks, what simialr processes are being used in both tasks? That would be useful to know, if the analog were to become part of an instructional sequence.

Is the methodology appropriate for the current state of knowledge about problem solving in chemistry in the domain of solution concentration? The authors rightly comment that interviewing students is time consuming, costly, and difficult to do with large sample sizes needed to ensure representativeness. Therefore, they opted for multiple choice tests which, unfortunately, do not yield much information on the precise causes of errors or the processes students used to solve problems. There is not yet much known about the problem solving enterprise by students, nor are the problems themselves well characterized. If the purpose of this research is to improve the teaching of problem solving in high school chemistry, would a smaller sample, interview or case study methodology have been more appropriate?

ELEMENTARY SCHOOL SCIENCE

Harty, H. and D. Beall. "Toward the Development of a Children's Science Measure." Journal of Research in Science Teaching, 21 (4): 425-436, 1984.

Descriptors--*Curiosity; Elementary Education; *Elementary School Science; Grade 5; *Measures Individuals; Science Education; *Sex Differences; *Test Reliability; *Test Validity

Expanded abstract and analysis prepared especially for I.S.E. by Paul L. Gardner, Monash University, Australia.

Purpose

The aim of the study was to develop a Likert-type instrument for measuring scientific curiosity in elementary school children.

Rationale

The authors argue that although curiosity is an important attribute which influences learning, relatively few empirical investigations have been conducted, and that this may be due to a lack of valid instruments.

They begin their paper with a succinct summary of the literature on the concept of curiosity. The operational definitions of Penney and McCann (1964) and Maw and Maw (1970) appear to lay the foundation for the authors' views of the nature of this construct. According to these definitions, curiosity is a behavioral attribute which involves approaching, exploring, manipulating, and reacting favorably to novel stimuli.

Research Design and Procedure

The CSCS was developed by means of an elaborate procedure, beginning with a large initial item pool. Many of the items were taken from other instruments, edited where necessary to give the item a science context. The pool was subsequently refined; six successive versions of the instrument were prepared.

In developing the instrument, the authors applied a wide range of psychometric procedures. Internal consistency reliability was assessed twice by means of Cronbach alpha. Item/scale correlations were calculated, and items with low values discarded or rewritten. Test-retest reliability was found using the fifth version. Concurrent validity was investigated by correlating the fifth version with scores on a set of discriminating items from the Penney and McCann (1964) Children's Reactive Curiosity Scale, and also with a single item in which students rated their own interest in science on a three-point scale. Predictive validity was studied by correlations with semester grades.

Face construct validity was investigated by asking a panel of two professors of elementary education and elementary teachers to criticize the first version of the instrument.

Content construct validity of the second version was checked with the help of an expert panel of judges, who were asked to rate the items on various dimensions.

The sixth (final) version of the CSCS consists of 30 items. Sample items:

1. Science magazines and stores are interesting
2. I like to watch television programs about science
3. I enjoy collecting leaves or other things from the outdoors

The response mode is the conventional five-point Likert scale (Strongly Agree, Agree, Uncertain, Disagree, Strongly Disagree) supplemented by smiling, unexpressive or frowning faces.

The final version was subjected to factor analysis with varimax rotation; the analysis yielded four factors which the authors labelled "novelty," "lack of clarity," "complexity of stimuli" and "surprise/bafflement." These descriptions appear to have been influenced by the writings of other researchers in the field of curiosity.

ABSTRACTOR'S ANALYSIS

Although I will be rather critical of some of the authors' arguments, my intention is not to denigrate their work. Quite the contrary: in many respects their use of a wide range of statistical procedures to develop a new instrument displays exemplary adherence to the principles of educational measurement. Other authors sometimes fail to clarify the conceptual basis of the constructs they are investigating or neglect to include the psychometric evidence which makes a detailed evaluation possible. Harty and Beall, in contrast, have presented their arguments and evidence in meticulous detail, in a form which makes criticism - and hence progress in the discipline - possible.

Many aspects of their paper are sound. Their description of the construct of curiosity has been tied to the literature on the subject; the large number of statistical procedures used in an attempt to evaluate the instrument, and the careful application of these procedures in order to refine the instrument are entirely commendable.

Four aspects of their paper, however, merit critical attention:

1. the construct validity of the CSCS is dubious, if one defines curiosity as something which is more than mere interest
2. the "smiling/frowning faces" response mode, adopted in order to reduce reading difficulties among some respondents, is potentially confusing when used in conjunction with unfavorable statements
3. there appears to be some confusion by the authors between the concepts of internal consistency and unidimensionality, and this leads them to misconstrue their factor analysis findings
4. the theoretical interpretation of some of their factors is unconvincing.

Construct validity

A central weakness of the paper is the mismatch between the authors' description of the construct they intend to measure, and the instrument they then produce to measure it. In my view, the instrument probably measures interest in science; I doubt if it measures curiosity. Obviously, my criticism hinges on the argument that the two constructs are not synonymous.

As already mentioned, the introductory sections of the paper present other authors' conceptions of curiosity. For example, Maw and Maw (1970) describe curiosity as reacting positively to unfamiliar elements in the environment; moving towards exploring and manipulating these elements; wanting to know more; scanning one's surroundings to seek new experiences; persistent examination or exploration of stimuli. Other writers use similar language.

A common feature of the various descriptions is that curiosity involves self-initiated action by the person towards some novel aspect of the environment. Obviously, a person who is curious about something is interested in it, but curiosity is more than just interest. Children can demonstrate interest by being willing to attend to some phenomenon that someone else has presented to them, but we would not describe a child as curious unless we had some extra evidence of self-initiated activity.

Peterson (1979), in her research on curiosity, operationalized the construct in a manner entirely consistent with this view. Students were ushered into a "waiting room" equipped with many types of scientific material prior to an interview which they thought was the purpose of the study. Their interactions with the material in the room were observed unobtrusively and used as behavioral indicators of curiosity.

Peterson's operational definition of curiosity in terms of sensori-motor behavior (approaching, manipulating and re-organizing objects) is consistent with Maw and Maw's conceptualization. It is doubtful, however, whether Harty and Beall's measurement technique displays the same consistency.

To begin with, there is the overwhelming problem of the validity of using a pencil and paper Likert scale containing verbal statements to measure a construct with a strong behavioral component. Even if it were possible to assess curiosity this way, it is doubtful whether most of the items in this instrument come close to doing so. Statements such as "I like to watch television programs about science", "Movies and pictures about volcanoes are interesting" and "I would like to listen to scientists talk about their jobs" are indicators of interest in science, and in fact are indistinguishable from items in other long-established measures of science interest: see, for example, the Meyer (1969) Test of Interests.

The issue here is not a mere semantic quibble over what to call a scale. As Peterson (1979) has shown, sensori-motor indicators of curiosity and verbal indicators of interest measure distinct constructs.

In the light of this analysis, some of the evidence concerning the validity of the CSCS is unconvincing. The finding that there were "significant differences between students who were interested in science and students uncertain about interest in science" (p. 425) is, of course, just what we would expect if the scale were just another interest measure. The evidence that the scale correlated strongly with the items taken from Penney and McCann (1964) Children's Reactive Curiosity Scale is hardly surprising either, since this scale contributed items to the initial pool used to construct the CSCS (see p. 427). And while judgments by elementary school teachers and professors (p. 430) are certainly helpful for establishing the appropriateness of a scale for use in schools, people without a substantial background in psychological measurement are unlikely to make critical comments on whether items accurately reflect a complex psychological construct.

Response Mode

In order to assist "students who have interpretation difficulties", Harty and Beall used drawings of faces to supplement the conventional five-point Likert Strongly Agree through to Strongly Disagree response mode (Two smiling faces = strongly agree, one smiling face = agree, expressionless face = uncertain, one frowning face = disagree, two frowning faces = strongly disagree). This is perfectly appropriate for favorable items, e.g. for the item "I like to watch magic shows". A student who likes magic shows will agree and associate this with a smiling face.

But what about unfavorable statements such as "It is boring to read about different kinds of animals"? Here confusion is possible. A student who disagrees with the statement - one who enjoys reading about animals - is meant to select the frowning face, because frowning symbolizes disagreement with the statement, not the emotion reflected in the statement. A young student who happens to like reading about animals and checks the smiling face to register that emotion might be forgiven! Obviously, if smiling/frowning faces are to be used in a scale for young children, only favorable statements should be included in the scale.

Internal consistency and unidimensionality

The authors report (p. 427) respectably high alpha values for two earlier, longer versions of the CSCS. Items in these versions with low item/total correlations were discarded or rewritten.

Later, they factor-analysed the final (30-item) version and found evidence of four varimax factors. In two places they appear to be confused about the interpretation of these psychometric procedures:

> Internal consistency reliability has been appropriately applied to sets of homogeneous items, that is, entities composed of equivalent units where items measure the same trait (curiosity) to about the same degreee. (p. 427)

And:

> The results of the factor analysis suggest that the 30 items might be used as a single scale to identify science-curious fifth grade youngsters. (p. 433)

This conclusion constitutes a straight-out misinterpretation of the findings. If the factor analysis yielded four varimax factors then this constitutes clear evidence that the CSCS is not unidimensional, that there is no single trait of curiosity underlying all the items in the instrument, and that scoring the CSCS as a single 30-item scale is not justifiable.

The problem in the argument here is that internal consistency (item homogeneity) is not synonymous with unidimensionality. Evidence that items correlate with the scale* does not provide compelling evidence that they correlate with each other (i.e., share a common trait).

Interpretation of the factor analysis evidence

The authors appear to have some difficulties in interpreting their factor analysis evidence. For a start, they claim (p. 433) that their four factors respectively account for 49.8%, 20.4%, 16.8% and

* Incidentally, item/rest-of-scale correlation is preferable to item/total scale correlation, since the latter statistic is spuriously inflated by the inclusion of the item under investigation in the total.

13.0% of the total variance of the instrument. Since these figures total 100%, this would imply that all the variance in the instrument is explicable, a most unlikely state of affairs. They undoubtedly mean that the factors account for the stated percentages of the explained variance, but do not report what proportion of the total variance has been explained.

The labels used to describe their factors are unconvincing, and it appears that their interpretations have been more heavily influenced by other researchers' findings than by the actual content of the CSCS items.

The simplest starting points for interpreting a factor are the pivotal items, i.e., the items having the highest loadings on the factor. Harty and Beall call their Factor I "novelty" and while this is not an inappropriate description, they fail to note that three of the four highest-loading items refer to space-travel.

Factor II is given the vague description "lack of clarity". A simpler interpretation is possible - the three highest loading items all refer to living things. It seems to be largely a nature-study factor.

Factor III was labelled "complexity of stimuli", another unconvincing description. The items with the highest loadings on this factor were:

15. I dislike to look at small objects through a magnifying glass.
29. It is boring to ask questions about how animals live.
6. I don't want to know how rainbows are formed.

The label "complexity of stimuli" for such items is rather puzzling. The science content referred to in these items is no more complex than that of items which load on other factors. What is noticeable, however, is that the three items listed are all unfavorable statements. In fact, of the eight unfavorable statements in the scale, six load on this factor. "Negative statements" would seem to be a more descriptive label. The emergence of this as a separate factor may indicate that liking is not necessarily opposite to disliking, a point first raised twenty years ago by Jordan (1965). However, it may also be linked to the response mode problem mentioned earlier: the factor may simply be distinguishing between students who interpreted and misinterpreted the instructions.

The argument presented here points to one of the dangers of using factor analysis as a post-hoc method of explaining the consistencies in a set of items. In my view, it is better to conceptualize one's dimensions first, allocate items to scales according to that conceptualization, and then use factor analysis to check the degree of fit between items and scales, and between conceptualized scales and actual factors. Since, at every stage of their scale development procedure, Harty and Beall seemed to have conceptualized "curiosity" as being unidimensional, one might then have expected them to be surprised by their four-factor varimax solution.

Over a decade ago, Peterson and Lowery (1972) argued that

> if a goal of science education is to preserve or nurture curiosity, then any assessment of that behavior must be based on a broad understanding of the nature of curiosity.

It would seem that those of us who are interested in developing such methods of assessment still have some work to do.

REFERENCES

Jordan, N. "The Asymmetry of 'Liking' and 'Disliking': A Phenomenon Meriting Further Reflection and Research." Public Opinion Quarterly, 29: 315-322, 1965.

Maw, W. H. and E. W. Maw. "Self-Concepts of High and Low Curiosity Boys." Child Development, 41: 123-129, 1970.

Meyer, G. A Test of Interests. Milton: Jacaranda Press, 1969.

Penney, R. K. and B. McCann. "The Children's Reactive Curiosity Scale." Psychological Reports, 15: 323-334, 1964.

Peterson, R. W. "Changes in Curiosity from Childhood to Adolescence." Journal of Research in Science Teaching, 16: 185-192, 1979.

Peterson, R. W. and L. F. Lowery. The Use of Motor Activity as an Index of Curiosity in Children." Journal of Research in Science Teaching, 9: 193-200, 1972.

Perrine, Walter G. "Teacher and Supervisory Perceptions of Elementary Science Supervision." Science Education, 68 (1): 3-9, 1984.

Descriptors--*Administrator Role; Administrator Attitudes; Elementary Education; *Elementary School Science; Science Education; *Science Supervision; *Supervisors; Teacher Attitudes.

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

Purpose

The purpose established for this study was to identify the perceptions of elementary teachers and science supervisors concerning current practices, and to describe the ideal process of supervision as perceived by the members of each group.

Rationale

What should the science supervisor do? Dowling (1977) argued that improving the science program by serving more as a resource person than an administrator was the most appropriate role. Others, of course, see administration as the most important function. If communication between supervisor and teacher is to be optimal, and teachers perceive administrators to be evaluators rather than equal partners in the educational process, communication between these two groups will be impaired. It thus becomes important to identify perceptions of science supervisors held by elementary teachers and the perceptions of elementary teachers held by science supervisors, because, with this knowledge, a program to eliminate communication problems could be designed. This study had as its purpose identifying these perceptions.

Research Design and Procedures

A 32 item questionnaire developed by the author to evaluate behaviors of supervisors of elementary science educators was used to survey the elementary science supervisors of New Jersey and a random sample of self-contained classroom teachers under their direction. The study was reported to include 29 supervisors from across the state and 470 randomly selected elementary (K-8) teachers. However, the investigator also indicated that in districts with fewer than 40 elementary teachers, all of them were asked to participate. Also reported was the fact that in the larger districts the teachers were selected randomly until approximately 20 viable returns were received. Teacher anonymity was maintained.

The investigator, in developing the instrument, first identified approximately 200 descriptive statements of supervisory practice. He then categorized this list into component areas and reduced it to 50 items. He converted the statements into questions and had them reviewed by a panel of supervisors and a panel of elementary school teachers. As a result of these exercises, the investigator assembled a 32 item instrument with a five point Likert-type scale. The instruments included the following components of supervision.

Component	Variation
Communication	Two way-one way
Working relationship	Assisting-directing
Leadership styles	Democratic-autocratic
Creativity/confidence	Positive-negative
Personnel position	Staff-line
Initiation structure	Positive-negative
Consideration	Positive-negative
Decision making	Group-autonomous

Teachers and their supervisors were requested to answer each question from two standpoints.

How does your science supervisor perform this task?

How should your science supervisor perform this task? Here is the example provided by the investigator

	Should	Does	
#13	()	()	develop the self-potential in a teacher?
#30	()	()	make clear his/her expectations of the teacher?

Supervisors were given the same questions and asked, "How do you perform your role?", and, "How should you perform your role?". I.e., should/do you develop the self-potential of the teacher? Demographic data were obtained from each individual participating and teachers were asked to write in comments about particular strengths they observed in their supervisor.

Construct validity of the instrument was established through using a jury of seven experienced elementary supervisors, "who agreed on the indexing of the items into aforementioned components." The factor analysis of the teacher's perceptions of both actual and ideal science supervisory behavior indicated only two major factors; instruction oriented and professional skills. These two factors accounted for 74% of the total variance for the actual behavior and for 80% of the total variance for the ideal behavior.

Cronbach's Alpha correlation coefficients used to demonstrate the instruments reliability were reported to be:

Teacher's ideal	0.69
Teacher's actual	0.83
Supervisors' ideal	0.49
Supervisors' actual	0.70

Findings

The author's two tables contained ratings of both teachers and supervisors and comparisons of their perceptions are presented as tables 1 and 2. The findings were:

1. The ratings of ideal supervisory behavior reported from the data collected from the teachers was reported to be significantly higher than were the ratings that supervisors indicated were ideal.

2. The ratings of ideal supervisory behavior reported from the data collected from the teachers were found to be significantly higher than the supervisors' perception of what ideal behavior was.

3. The differences between the perceptions of actual behavior between teachers and supervisors were not significant.

4. Similarly, the differences between ideal and actual perceived by teachers and supervisors were not significant.

Correlation coefficients between 15 demographic variables and the teachers' perceptions of supervisors were also calculated. The variables studied were: sex, grade level teaching, teacher's age, minutes per week of science, teaching years in the system, total years in teaching, the number of graduate credits taken, science methods courses taken, science content courses taken, inservice courses taken in the past five years, teacher interest in taking a science course, single text versus multi-text used, the particular text used, and teacher's subject matter preference. The authors reported that there were several significant correlations. However, because none of them exceeded 0.12, they were not reported.

Supervisory strengths reported by approximately 11% of the 432 participating teachers included:

1. Subject expertise: Knowledgeable, competent and experienced.

2. Human respect and understanding: Encouraging, cooperative, supporting, assists, flexible, creates confidence, communicates well.

3. General positive comments: Great, helpful, good, excellent, terrific, etc.

Supervisory strengths reported by appoximately 7% of the 432 participating teachers included:

1. Supportive supervisory techniques: Accessible, available, see regularly, attentive, dependable, interested, and responds.

2. Positive personality traits: Pleasant, compatible, even tempered, good rapport, and friendly.

3. Attitude: Enthusiastic, very involved, concerned, great commitment, dynamic, generates enthusiasm, etc.

Supervisory strengths reported by approximately 4% of the 432 participating teachers included:

1. Supplier of: Materials, suggestions, programs, information, etc.

2. Organized: Has it all together, systematized, efficient, etc.

Interpretations

Teachers held higher expectations of the supervisor than did the supervisors. Both groups rated the supervisor's present behavior to be significantly less than ideal behavior. This difference, the authors suggested, indicates that the supervisor's job description needs to be clarified. The authors concluded that only two of the supervisory characteristics were significant. These were the ability of the supervisor to provide technical assistance, and ability of the supervisor to be supportive in general reinforcing. All of the author's recommendations related directly to these two conclusions.

ABSTRACTOR'S ANALYSIS

Perrine attempted to identify the perceptions of elementary school teachers and supervisors concerning the actual practices of science supervisors and what each group considered to be ideal practices. As could be anticipated, perceptions differed. It was not at all unexpected to discover that teachers thought that the supervisors should be doing more, and that the supervisors felt that they were not getting done as much as they would like to accomplish. Perrine attempted to ascertain both the teachers' and the supervisors' perceptions of the ideal role of the supervisor, and the actual role that the supervisors were playing.

Perrine used experienced science supervisors as a jury to establish the validity of the instrument used in the study, and later calculated Cronbach's alpha to determine the instrument's reliability. Both of these standard steps yielded positive indicators of the instrument's validity. Sample items were provided as is customary, and generally allowed by the publishing journal. This writer would have appreciated the opportunity to study the entire instrument as it is most difficult to discuss the results of a perception study when one does not know precisely what each item was. Journal editors need to be encouraged to encourage authors to include more sample items or, when possible, the entire instrument they are using in their study.

The research design, as described, was a bit "fuzzy." Random is random and infinite. Random ended, or seemed to end, or was bent a bit. Yet, it seemed very logical to proceed as the author proceeded. "Random" is a magic word and most researchers strive for it at great expense. Yet, in many instances, random could be less desirable than a purposeful sample. This researcher, it seems, used a purposeful sample, or at least he used a sample that could be so labeled.

The instrument consisted of 32 questions, a Likert-like scale, and the opportunity to respond to both Should and Does. In essence, it was a 64 item instrument because two Likert-like responses were available for each question. In addition, the teachers were asked to provide free responses on their perceptions of the supervisor's strengths. Perrine did little with the findings from the open response question because he discovered that the significant correlations were all low (0.12 or less). His was an appropriate but conservative action. In reviewing his findings it was obvious that one could conclude that having useful information was the most important characteristic of a science supervisor. The next five important strengths of the science supervisor reported by Perrine could all be related to interpersonal relationships. While finding one such indicator may not be significant, a review of his data might allow the conclusion that finding five very similar strengths could be very significant. This writer thinks that what was discovered was that there were three significant supervisory behaviors. Presented in order of their importance they are:

1. Being knowledgeable
2. Being able to communicate effectively
3. Being organized.

However, while being knowledgeable was cited as #1, the next five findings all dealt with communication. A trend? While one finding would not suggest much, five very similar findings, all in the same direction, are much more important, and probably much more significant than one finding. This could have been pursued further.

The free response part of the author's instrument may have yielded the most important information of the study. All else was too easy to anticipate. (However, this writer did not see the entire instrument.) While "knowing it" turned out to be number one as has often been the case, the next five findings had to do with "communicating it." It is this writer's opinion that "knowing it" is most important, UNLESS THE KNOWER can't communicate it to someone else. Progress is not made by talking to oneself.

This report, like many others, emphasizes how important it is to know something. But, knowing something can be unimportant if the knowing can not be communicated. This has very important implications for the training of both teachers and supervisors.

REFERENCES

Brewer, A. C. "The Role of the Science Supervisor." In M. B. Harbeck (Ed.), Handbook for Science Supervisors, Washington, D. C.: National Science Supervisor Association, 1967.

Dowling, K. W. "Hand Holding in Science Education." National Science Supervisors Association Newsletter, 14, 1977.

Harris, B. M. Supervisory Behavior in Education. Englewood Cliffs, NJ: Prentice-Hall, 1963.

Sternig, J. "The Elementary Grades in the Nation's Schools." National Schools, 65, 1960.

Whitla, D. K. "Essentially Elementary Science: A Report on the Status of Elementary Science in Massachusetts Schools." Cambridge, MA: Harvard Graduate School of Education, 1973.

SCIENCE PROCESSES

Tobin, Kenneth. "Student Task Involvement and Achievement in Process-Oriented Science Activities." Science Education, 70 (1): 61-72, 1986.

Descriptors--*Achievement Gains; Intermediate Grades; Junior High Schools; *Process Education; *Science Activities; *Science Education; *Student Behavior; *Time On Task

Expanded abstract and analysis prepared especially for I.S.E. by Steve Tipps, Midwestern State University, Wichita Falls.

Purpose

The purpose of this study was to describe the engagement behavior of students during process science lessons and relate student engagement to achievement of process skills.

Rationale

Although process learning and engagement time have been topics of much recent educational research, no research has been undertaken which investigates how time on different activities might be related to process learning. Students may be engaged in either overt-active or covert-reflective behavior. Assessment of formal reasoning is necessary, also, as formal reasoning is related to science achievement.

Research Design and Procedure

Eight lesson plans were provided for teachers of 15 intact classes in grades 6 and 7. Each lesson plan included three parts: planning, data collection, and data processing (organization and interpretation). The lessons were designed to encourage student engagement.

Twelve students randomly chosen from each class were observed during four of the eight science lessons. Engagement was observed by a trained rater using a coding system with eight engagement categories. The predominant student activity for a four second (per minute) observation was recorded. Teacher behavior for each minute of observation was also recorded. Field notes were also taken to provide qualitative descriptions of the classes. The Test of Logical Thinking was used to assess formal reasoning, and a modified form of the Test of Integrated Science Processes assessed process achievement.

Findings

Teachers allocated 25% of lesson time to planning investigations, 37% to data collection, and 30% to data processing. The average proportion of time on task was very high (.88 to .94) for the 142 students in the final sample. However, field notes showed that overt engagement during planning and processing was typically limited to a few students answering and volunteering.

Engagement was primarily covert in the planning and processing parts of the lesson and active only during the data collection. Of the eight categories of behavior observed, only three or four were large enough to enable meaningful analysis. The means for "attending" and "collecting" were .37 and .31 and accounted for over two thirds of the total engagement time per minute. Over the four lessons, overt engagement increased linearly from 39% to 49%, while covert engagement decreased from 42% to 32%.

A relationship between formal reasoning and engagement behavior was found in two instances. Formal thinkers tended to engage in slightly more generalizing behavior ($r = .17$, $p < .02$), while students with lower scores were more likely to be engaged in comprehending tasks ($r = -.28$, $p < .001$).

Formal reasoning ability accounted for about 36% of the variation in science process achievement. Overt collecting and planning were also significantly related to achievement. Engagement in comprehending was related to lower scores in process achievement ($r = -.20$, $p < .0$[illegible]) Seventh grade achievement was higher than grade six achievement. "A five-variable model which included reasoning ability, grade level, collecting, planning, and attending, was found to account for 47% of the variance in process skill achievement." 42

Interpretations

The researcher concluded that, even in lessons designed to engage students in process activities, many students were not actively engaged in the planning and interpretation phases. Although the students were judged on task, 37% of their time was in attending behavior such as watching, listening, etc. Whole class grouping was usual during planning and interpretation phases in which relatively few students were active. The increase in engagement through the sequence of lessons, however, offered encouragement that students could become more active as they become more skilled. Small-group or individual work in experiments are suggested as ways to further increase active involvement.

The researcher also suggests that too little time may have been spent on the lessons to master the processes. "More than eight lessons..." or "a higher proportion of instructional time" may be necessary to develop process skills. Several alternative explanations were offered for the relationships between engagement in different tasks and achievement. The test emphasized planning and interpretation and may have not have been balanced in measuring data collection skill. Skill in data collection may have contributed to achievement in these areas without showing up on its own. Higher achieving students may already engage in planning and interpretation. Reducing engagement of lower achieving students in comprehending and increasing their engagement in planning might bring about improved achievement.

Formal reasoning ability was found to be a strong determinant of process skill achievement. Consideration of reasoning ability in development of skills continues to be necessary. Covert student engagement did not contribute to achievement; however, no attempt was made to manipulate classroom characteristics such as wait time.

Generalizing behavior was consistently low among students. The researcher suggests a threshold of generalizing which may be necessary to demonstrate achievement. Since so few overt behaviors were identified during the planning and interpretation phase, modification of lessons to encourage overt activity is recommended. If more behaviors were represented, a better picture might be obtained regarding engagement and process achievement.

ABSTRACTOR'S ANALYSIS

Tobin's study is a continuation of research on science process learning, classroom activity, and formal reasoning which has been going on for the last decade by Tobin and others. This body of research continues to be a major contribution to science education. In this study, a direct connection is sought between what students actually do and what they learn about science processes. Careful data collection is the core of the research, however, the use of qualitative methods in the form of field notes is a worthwhile and important aspect. Only when the data are put into context can they be interpreted meaningfully.

The study must be considered exploratory. As thorough as the lesson design and data collection were, a second effort at codifying the classroom would certainly benefit from the problems identified in this study. The lessons failed to elicit overt behaviors in the quantity needed for analysis. Covert attending was not very fruitful in tracking the precursors to achievement. Manipulation of classroom techniques for greater overt engagement might enable researchers to assess learning more easily. However, covert engagement (thinking, reasoning, considering, being puzzled) is also important; it would be unreasonable to replace all covert behavior with observable behaviors.

The random sampling of students from intact classes appears to be a convenient, yet adequate way of selecting subjects. The researcher is cautious in generalizing and calls for replication with improvements in the lessons.

Interpretation of the data is not easy. In addition to the failure to observe several of the engagement behaviors at any frequency, the relationships among lesson, engagement, achievement, and cognitive ability are very complicated. Does greater ability influence type of engagement? Would that ability result in higher achievement regardless of the lesson? Is the test overly weighted on planning and processing without due consideration for data collection? These are complex questions, and Tobin has some understandable difficulty explicating them. The analysis and interpretation sections are not as well organized or stated as a reader might desire. For example, discussion of "target students," or students who were more engaged with the teacher during planning and data processing, is not

44

always clear. The reference to one table is wrong in the text, and one table is never discussed. The units of measurement on several variables are puzzling. Another draft of these sections might have helped.

One issue not mentioned in the research is that of prior process skills. Without a pre-test, it is difficult to tell whether the lessons or the engagement had any impact on achievement. If the higher achieving students already had the skills, they probably would have been more active in class. However, the students who had few or no process skills might have actually gained a great deal through attending and interacting. The gradual increase in student engagement over the lessons is encouraging. Tobin did not report whether the engagement for high versus low (achievement or reasoning) students changed over the lessons. This would appear to be an interesting question to pursue.

45

Berger, Carl F. "Attainment of Skills in Using Science Processes: I. Instrumentation, Methodology and Analysis." Journal of Research in Science Teaching, 19(3): 249-260, 1982.

Descriptors-Adolescents; *Computer Oriented Programs; *Elementary School Science; Elementary Secondary Education; *Microcomputers; Models; *Process Education; *Research Methodology; Science Education; Secondary School Science; *Skill Development

Expanded abstract and analysis prepared especially for I.S.E. by Rodney L. Doran, State University of New York at Buffalo; Michael Dryden, Arizona State University; and Guy Ilogu, University of Lagos, Nigeria.

Purpose

As stated by the author, the purposes of this study were to (1) develop instrumentation, methodology and analysis techniques to measure attainment of process skills in "estimation," and (2) observe the effect of the amount and kind of information on the attainment of these skills of estimation of linear distance.

Rationale

The researcher stated that a major goal of science curricula in the late 1960's was the attainment of process skills, yet little evidence has been gathered on the short term acquisition of these skills. In a brief analysis of the research, he reiterated the finding of the second report of An Evaluation Model and Its Application (Walbesser et. al., 1968). The report showed an anomolous decrease in linear estimation skills from first to second grade. This was the researcher's basis for using linear estimation as the main construct to be investigated.

The researcher capitalized on the opportunity to use microcomputers for their potential use of on-line data gathering and storage. It also had the potential for presenting data in a fully controlled fashion. The DART program was adopted for use on three then widely-used microcomputers: the PET, the TRS-80 and Apple II.

Research Design and Procedures

The author states that he rejected a time series approach for a multiple regression approach that measures variation in results while accounting for individual variation. The sample population was 10 adolescents, ranging in age from 10 to 14. The author claimed that their academic abilities were within the "normal" range. Information as to gender of the adolescents was not provided.

The experiment was as follows: a vertical wall was portrayed on the right hand side of the monitor. To this wall was attached a balloon (spot). The base of the wall was marked 0 and the top 100. Units were not provided. Each subject was asked to make several attempts to determine the exact position of the balloon. Upon successful completion of an estimate, a new point was provided. Ten estimates were scheduled for each of the three levels. The amount of information supplied about the status of previous estimates and further clues determined the "level" of the task. Level 1 had considerable information; level 2 had less; and level 3 had the least information provided. Students progressed through the three levels (from 1 to 3) with a researcher entering an estimate on a microcomputer keyboard. The researcher anticipated that three general strategies would be followed. The first used prior estimation results to narrow the uncertainty (ladder approach). The second approach was similar, but used alternating high and low estimates to narrow the uncertainty of estimation (bracket approach). The third approach was random guessing. It neither reduced the uncertainty of estimation nor did it use prior estimation results. The basic premise of the experiment was that performance would be proportional to the amount and kind of information supplied. Further, within each level, performance would improve with time (practice).

Some major assumptions of the study were that an understanding of a length and number was being used and that there was an association between performance and the visual cues provided. The criterion for success, measured as the average time of estimation across the set of problems at each skill level, was taken to be a measure of the transfer of estimation skill.

Students were given two days rest between each level of the experiment. The amount of time required to "estimate" at each level ranged from 10 to 23 minutes. For each student the experimental data (number of estimates per problem and time per estimate) were stored for later analysis.

Findings

For the sample of 10 students, a total of 856 "data points" were gathered. A mean estimate per problem was 1.4 with a standard deviation of 2.2. The range of estimates per problem was from 1 to 5.

The researcher's unit of analysis seemed to be the problem number. With this information, he used multiple regression as the analysis technique. As far as strategy was concerned, the following approaches were inferred from the data:

a. 1 problem via the random approach
b. 8 problems were solved using the "bracket approach"
c. 18 problems via the ladder approach.

The average time per trial per problem was the criterion and the subject and trend vectors were used as the predictor variables. These variables were not well defined. Results showed that student performance improved only within the first two difficulty levels. The investigator inferred a "transfer effect" from this information.

In terms of student use of the several strategies, the investigator concluded that five students used the ladder approach, two the bracket approach, two students alternated between the ladder and bracket approaches and one student progressed from a random to the ladder approach. The average time of estimates for the 10 problems range from 1.5 to 17 seconds.

From these indings, the researcher "concluded that microcomputers were a valuable instrument for gathering and recording data." "The model using regression analysis was an effective tool to study estimation." He also concluded students used effective strategies and that learning occurred.

ABSTRACTORS' ANALYSIS

The researcher was attempting to perform too many tasks within one study. Validating instruments, methodologies or analysis techniques are each worthy of their own study. Validation of the techniques was never performed. The major goal of measuring change in estimation procedures seemed to take a back seat to the vehicle of instrumentation, the microcomputer. He stated that the study is between a large "macro study" and an individual based "micro study". However, with 10 students, it may have been wise to describe the investigation as a case study. Using the results of a study based on first and second graders provided a weak criteria for selecting the process of linear estimation with 12 to 14 year olds.

Indeed, based on the researcher's description, two-dimensional estimation of horizontal extensions seems to be what was being measured, not linear estimation. Control for student characteristics, such as field dependence/independence, locus of control and gender were not discussed.

The graphical learning models were useful in explaining anticipated results to the reader. However, the lack of difficulty within the problems raises the question of a "ceiling effect." This confounds the issue as it implies that the reduction in average time per estimate may not be due to the retention of prior information but an adjustment in the microcomputer environment.

The use of multiple regression with time of estimation as the criterion and individual students with trends as the predictor variables is difficult to interpret. The criterion was very confusing. The number of estimates per problem was initially abandoned in favor of time of estimation. Yet, in his conclusion, he states that the time of estimation was not a good criterion of the process of estimation. Despite this contradiction, there were other ill defined variables. As stated earlier, to predict average time of estimation the author used a student vector and a trend vector. It appears as if each student was treated as separate treatments with the problem number being different cases. In this way he could measure student variation between problems. However, by using the average time of estimate, the author eliminated any change to evaluate within student variation between problems. With this design, the R squared values would be measuring the amount of variation in the average time of estimation per problem as explained by individual students and by the trend vector. Further confusion occurred in that his equations show the student vector as an independent variable but his graphs show the problem number as the independent variable. Is the student the independent variable and the problem number the case or vice versa?

The abstractors could not determine which trend was being measured since it looked like three different runs by skill level were attempted without any measure of interaction. In trend analysis, the independent measure must be continuous. Also, the measurement of the variable should have a substantially high reliability. Reliabilities were never reported and there were no continuous variables. Assuming that a valid trend vector was analyzed, this method could be powerful in exploring the non-linear nature of a relationship.

In the linear least squares regression fit, the author illustrates almost perfect linearity yet to increase the R squared value, a third order correlation was analyzed. Why? One conclusion drawn was the "practice...in level III was sufficient to prevent further improvement of average estimation time." In other words, by going through levels I and II the students mastered the DART game. He went on to show this by using problem one in the three levels as a repeated measures design.

Basically, the article seems to be an attempt to explore new methods of research and evaluation. Unfortunately, it does not stand up to the rigor of scientific scrutiny through validation of constructs. The sample was much too small for generalizations to be made and there was a high probability for capitalization of chance with such a small sample. The author had more analysis techniques than he had students.

CURRICULUM AND INSTRUCTION

51

Renner, John W., Michael R. Abraham, and Howard H. Birnie. "Secondary School Students' Beliefs About the Physics Laboratory." Science Education, 69(5): 649-663, 1985.

Descriptors-Enrollment; Interviews; Laboratory Procedures; *Learning Processes; *Physics; Science Education; Science Instruction; *Science Laboratories; Secondary Education; *Secondary School Science; *Student Attitudes

Expanded abstract and analysis prepared especially for I.S.E. by John R. Staver, University of Illinois at Chicago.

Purpose

The authors' purpose is to describe a successful effort to increase the high school physics enrollment in Norman, Oklahoma, and to further delineate the role of the physics laboratory in this endeavor.

Rationale

The authors argue for an alternative hypothesis to the supposition made by the National Commission on Excellence in Education concerning decreases in student enrollments in academic courses. The Commission's hypothesis is that students avoid challenging academic courses because they are difficult. When choice is possible, according to the Commission, students elect to take easier alternatives. The Commission's solution is to make presently elective science courses required. The authors argue that this is short-sighted. Rather, they suggest that students are like most people in that they avoid hard work when they don't see any payoff. The way to attract students into elective physics and chemistry, according to the authors, is to teach these courses in a manner that allows students to see a payoff.

The authors assume that the nature of the entire discipline of science is inherently interesting to students. They go on to suggest that students may avoid high school physics because physics instruction does not reflect the nature of physics as science. The typical instructional strategy in physics is inform-verify-practice. A teacher or textbook informs students about a topic. Next the students have an experience, often in the laboratory, in which they verify the truth of what they were told. Finally, students get to practice with the new knowledge by doing problems.

The authors maintain that the inform-verify-practice strategy concentrates only on the products (knowledge) that physics has produced. It is devoid, they argue, of any need for students to coordinate learning experiences into a logical framework. It is devoid of doing physics! Instead, the authors state, students should have opportunities in the laboratory to collect data on a phenomenon, then interpret it by developing a logical system. Such instruction requires: 1) that students remain uninformed about a concept before they experience its phenomena in the laboratory; 2) that the experimental directions not inform students about the concept, 3) that the experiment's directions, however, allow students to collect data which they can interpret, thereby identifying and comprehending the concept; and 4) that the experiment's plan present students with opportunities to apply the concept after they have identified it.

The authors offer the well-known learning cycle approach (Atkin and Karplus, 1962) as an instructional model consistent with their rationale. The three phases of the original SCIS learning cycle, exploration-invention-discovery, have been renamed exploration-concept introduction-concept expansion, but the contents remain as before.

Some impressive case study numbers are provided in support of the authors' claims. Between the 1973-74 and 1982-83 school years, the physics and chemistry enrollments in the junior and senior classes of the Norman, Oklahoma, Senior High School rose from 3.8% to 15.8% and from 11.5% to 24.4%, respectively. Moreover, the higher percentages have remained steady for about five years. Yet, physics and chemistry remain challenging elective courses. Why did the enrollments increase?

The authors discount reasons such as: 1) requiring physics and chemistry; 2) recruiting students by faculty or administrators; 3) new stimuli at the local, state or national level that would motivate students to take physics and chemistry; or 4) student awareness of new career goals that would require these courses. Rather, the increases resulted, in the authors' opinion, because students who enrolled in physics and chemistry told other students that these were good courses, that they should take physics and chemistry. Why did students transmit such positive reviews of difficult subjects? The authors argue that the single common thread is the learning cycle approach. Each physics and chemistry teacher (there have been twelve who occupied the single physics and 2.5 chemistry teaching positions during the time period) has used the learning cycle approach.

It was therefore hypothesized that the enrollment increases are due largely to the positive attitudes of students and that these attitudes are themselves due to the implementation of the learning cycle approach, particularly to the crucial role of the laboratory in the learning cycle.

Research Design and Procedure

The authors selected a naturalistic (Welch, 1981) design and used interviews and case studies to test their hypothesis. The instructional context within which all interviews were conducted is a physics curriculum comprised of thirty-six learning cycles. Sixty-five (65) seniors enrolled in one of three physics classes taught by a single teacher composed the population. Students were randomly chosen from this population for interviews, but any students who expressed reluctance to be interviewed were not forced to participate. This resulted in 23 participants, 9 females and 15 males (sic). These participants were interviewed in a private room about: 1) their knowledge of physics; 2) possible changes in their understanding of physics, and 3) their feelings about the learning activities, particularly the laboratory. All interviews were audio recorded, then identified portions of the tapes were transcribed for analysis. Three criteria were used to pick out tape segments for analysis: 1) the student had to use the phrase "the laboratory" in response to specific information requested by the interviewer; 2) the student's reference to the laboratory was direct, not a high-reference "I thought the student meant" by the interviewer; and 3) "The laboratory" meant specifically those exploration and idea expansion activities that took place during the learning cycle approach. "The laboratory" specifically does not refer to verification experiences of students.

Findings

A summary of the authors' findings is given below:

1. 21 of 23 students interviewed reported very positive feelings about physics laboratory experiences. They felt that the physics laboratory, as used in the learning cycle, was a positive influence on their learning.
2. Two students expressed negative feelings specifically about the role of physics laboratories in the learning cycle approach. No negative feelings were expressed about physics laboratories in general.

Interpretations

The conclusions, inferences, and implications set forth by the authors focus on the role of the laboratory. They are summarized below:

1. Students prefer laboratory activities because such experiences:
   - A. help students remember;
   - B. encourage thinking about observed phenomena;
   - C. are less confusing, more interesting, and more concrete than other instructional formats;
   - D. make concepts more believable and understandable.
2. Use of the laboratory to introduce a concept, which is then followed by a discussion (the learning cycle approach) is the key to positive feelings.
3. Physics teachers must consider the laboratory as <u>essential</u> to teach physics as science, not as the products of science.

4. Students enjoy learning physics through the learning cycle approach and prefer laboratory based knowledge acquisition.
5. The above mentioned inclinations cause students to recommend physics to others via the "grapevine."

ABSTRACTOR'S ANALYSIS

Few research studies provide directions for practice and future scholarly study as this one does. I begin my analysis not with a critique of the strengths and shortcomings of the investigation, not with a summary of how these results fit into the existing knowledge, but with an analogy because I think that full comprehension of the contribution of this work is facilitated best by such a beginning.

In science methods classes I often ask students how they would get a cat to come out of its hiding place. Moreover, I ask them to base their suggestions on particular characteristics of cats. Adjectives such as curious, ornery, independent, loving, stubborn, and affectionate, are among those frequently mentioned, and two methods represent the extremes of a wide range of suggestions. One way is to reach in, grab the nearest appendage, and pull the cat out. Cats usually react negatively to this approach, and quite often first aid is needed to treat scratch and bite marks. Another way is to drag a long string in front of the hiding place. Within a few seconds, the cat appears, curious about what it sees. I then suggest that students are similar to cats in several dimensions, particularly in the adjectives that we use to describe them. Like cats, our students are also intelligent and capable of choosing what they like and don't like. I personally lament the fact that teachers often forget this. We make, all too many times, the invalid assumption that our students love the subject matter we teach (in this case physics) as much as we do. We forget an ancient but still valid Tyler dictum, "... the value of beginning with present student interests as the point of departure" (1949, p. 11). The authors have provided some compelling evidence about what happens when physics becomes interesting to students.

This research offers an important object lesson to teachers and researchers alike. The lesson for teachers is to take advantage of the old Tyler (1949) dictum, to design science so that it interests and challenges students, so that they can see value and payoff. Students then will not only demonstrate how capable they really are, but also they will recruit others through the student "grapevine." It is one thing to have a counselor tell a student to take a tough science elective because it is good preparation for college. That's like telling someone to take castor oil. Its taste is revolting, but it's good for you. It is quite another to have students telling other students that, yes, this elective course is tough, but it is also interesting and it has a payoff. I prefer that students recruit other students.

The lesson for researchers lies in the source of these data and the design of the study. The authors asked a most important question, chose (in my view) a proper design, and clearly interpreted their results. The lesson for fellow researchers is to replicate, to cross check, to extend these results in more quantitative ways. The findings of naturalistic research should always be treated as questions, as alternative hypotheses to be tested quantitatively. The physics taught as science develops positive student attitudes which in turn causes students to recommend physics to other students is certainly an hypothesis that deserves further study.

The authors contribute significantly to the literature by providing yet another piece of evidence concerning the effectiveness of the learning cycle as a curriculum and instructional model. John Renner, with several colleagues, has devoted a large portion of his scholarly career to this endeavor. In previous research, it has been shown that: 1) students who studied science via the learning cycle (in this case SCIS Material Objects) increased their conservation reasoning over students who studied science from a textbook (Stafford and Renner, 1971); 2) that when SCIS Material Objects was used as a reading readiness program, it increased significantly the reading readiness of students (Renner, Stafford, Coffia, Kellogg, and Weber, 1973); 3) that the learning cycle encourages the intellectual development of concrete reasoning students (Purser and Renner, 1983); and 4) that the learning cycle is an effective method for teaching concrete concepts to concrete students (Purser and Renner, 1983).

Recently, Lawson (1986) has also stressed the importance of the learning cycle in his delineation of a theory of instruction for concept acquisition and reasoning development. The learning cycle plays an integral role, in that it provides students with opportunities to become aware of misconceptions, to argue and test them, and to reconstruct or develop a more adequate concept and reasoning pattern.

I close my analysis with a final thought. I often state in methods class that every student deserves a chance to think each day. The implications of this study, the existing body of knowledge, and Lawson's (1986) theory are clear. An excellent vehicle already exists, one that is effective for concept acquisition, for reasoning development, for getting students to think. It is the learning cycle, and it is our job as researchers to continue testing its worth, as science teacher educators to teach our students to use it effectively, and as professional development consultants to help all science teachers become aware of and use it.

REFERENCES

Atkin, R. M. and R. Karplus. "Discovery or Invention?" The Science Teacher, 29(5): 37-40, September, 1962.

Lawson, A. E. "Integrating Research on Misconceptions, Reasoning Patterns and Three Types of Learning Cycles." A paper presented at a U.S.-Japan seminar on science education, East-West Center, University of Hawaii, Honolulu, Hawaii, September 14-20, 1986.

Purser, R. K. and J. W. Renner. "Results of Two Tenth-Grade Biology Teaching Procedures." Science Education, 67(1): 85-98, 1983.

Renner, J. W., D. G. Stafford, W. J. Coffia, D. H. Kellogg, and M. C. Weber. "An Evaluation of the Science Curriculum Improvement Study." School Science and Mathematics, 73(4): 291-318, 1973.

Stafford, D. G. and J. W. Renner. "SCIS Helps the First Grader Use Logic in Problem Solving." School Science and Mathematics, 71(2): 159-164, 1971.

Tyler, R W. Basic Principles of Curriculum and Instruction. Chicago: University of Chicago Press, 1949.

Welch, W. W. "Naturalistic Inquiry and Conventional Inquiry: An Evaluation General Session Address." A paper presented at the annual meeting of the National Association for Research in Science Teaching, Grossingers, New York, April 5-8, 1981.

Volk, T. L., H. R. Hungerford and A. N. Tomera. "A National Survey of Curriculum Needs as Perceived by Professional Environmental Educators." Journal of Environmental Education, 16 (1): 10-19, Fall, 1984.

Descriptors--*Curriculum Development; *Educational Needs; *Educational Objectives; Elementary Secondary Education; *Environmental Education; Higher Education; National Surveys; *Needs Assessment; *Teacher Education

Expanded abstract and analysis prepared especially for I.S.E. by Uri Zoller, University of Haifa-Oranim, Israel.

Purpose

The stated purpose of this study was to assess the curricular needs in environmental education (EE) at all school levels (K-16) in the United States as perceived by professional environmental educators.

Specifically:

- The current status and desired state of EE curricula.
- The need for new curricula.
- The anticipated use of new curricula by teachers.
- The need for inservice teacher training concerning new EE curricula.

Rationale

Previous studies have indicated that existing EE curricula - nationally and internationally - fall short of meeting the major objective of EE that is, the "production" of an environmentally literate citizenry: individuals who are not only aware of the environmental consequences of the recent unprecedented technological progress, but also have developed the attitudes, skills, capacity, and will for a citizenship participation in environmental problem-solving.

An assessment of existing weaknesses and needs in EE, relative to a stated goal criteria, is thus required in order to strengthen on-going programs or to develop and implement new goal-oriented ones which will respond appropriately and effectively to the assessed needs. The underlying assumption is that the needs can be assessed by surveying the perceptions of professional educators and action should be taken accordingly thereafter.

Research Design and Procedure

A definitive set of EE goal levels and associated goal statements adapted from "Goals for Curriculum Development in EE" (Hungerford et al, 1980), which was subjected to rigorous validation by a jury of nationally recognized environmental educators, has been used as criteria against which the variables included in the study were assessed.

The self-administered questionnaire (EECNAQ) thus developed contained 15 goal statements. The 169 randomly selected U.S. professional educators (from the membership of the National Association for Environmental Education and Conservation Education Association) to whom the questionnaire was direct-mailed were asked to answer the following five questions concerning each of the goal statements (a Likert-type five response category + six options for not responding), relating to each of elementary, middle school/junior high, secondary, college/university levels (i.e., a 15 x 5 x 4 format):

1. To what extent is this goal important?
2. To what extent do existing curricula accomplish this goal?
3. To what extent is there need for new curricula addressing the goal?
4. To what extent would new curricula addressing this goal be used by teachers?
5. To what extent would inservice teacher education be needed for new curricula addressing this goal?

Findings

Based on the computation of the mean response to each item, the following findings were reported.

1. The EE goals are perceived as being important at each academic level; and importance increases with each level (all mean responses fell within the range of moderately-considerably-completely important).
2. Although perceived accomplishment of EE goals increases somewhat across the academic levels, it rarely exceeds the moderate extent of accomplishment. Thus, at each academic level, there exists a gap between perceived goal importance and perceived goal accomplishment, particularly with respect to high level goals (evaluation of issues and citizenship action).
3. Professional educators believe that new curricula in EE are needed particularly at the secondary level followed by the college, middle school, and elementary levels.
4. Teachers are anticipated to use new EE curricula to a moderate extent. Those addressing awareness level goals would gain greater use at the elementary and middle school levels, whereas, those addressing evaluation and citizenship goals would be used to a greater extent at the secondary and college levels.
5. Inservice teacher training is perceived as needed (within the range of the moderate-considerable categories) in the following order of priorities: secondary, middle school, elementary, college levels.
6. Except for the college level, the perceived need for inservice teacher education exceeded the perceived need for new curricula.

Interpretations

The authors interpret the results of their study in terms of a considerable discrepancy perceived between the desired state of EE and its existing status. This discrepancy would imply definite needs within EE in the U.S. They conclude that the results of their study provide a clear mandate for increasing inservice teacher education and curriculum development in the field of EE (focusing on and addressing its goals) at all academic levels. Also, they suggest that appropriate inservice teacher education in the use of newly-developed curricula will increase the use of these programs.

They conclude by stating that there is an immediate and critical need both for the development of new curricula which address all the goal levels of EE at all academic levels, and for inservice goal-oriented teacher education for all goal levels.

ABSTRACTOR'S ANALYSIS

Although the purpose of the present study was stated explicitly as "to assess perceived environmental education curriculum needs in the United Stated (K-16)," the implicit purpose, which is rather obvious right from the outset (and which became explicit at the conclusion), was to back by objective research findings a mandate for the development of new curricula and the increase of inservice teacher education (e.g., implementation of new programs) in the field of EE at all levels. In view of the felt needs shared by many involved in environmental education, it is no wonder that the implicit purpose played a major role in the conceptualization of the problem studied. We are dealing, therefore, with a "research for action" type; that is, research aimed at providing an objective data basis for decision- and policy-making in education systems.

Consequently, it appears appropriate to analyze the study on two different levels: the first - its adequacy as far as its internal validity is concerned; namely, the appropriateness of the conceptual framework, methodology, and design within the study perse, and the validity of the interpretations, conclusions, and generalizations derived.

The second level should address the external validity issue; that is, to what extent a subjective assessment of needs is scientifically, objectively adequate for decision on policy-making purposes, particularly when the subjective perceptions elicited are from those who have "major stakes" in the outcomes.

Assessments of opinions, perceptions, and attitudes constitute a common practice in educational research, science education and environmental education included. Clearly, the perceptions of the professional educator community concerning the current status and the desired state of EE are crucial, since the "general consensus" in this respect is a major factor in determining the future status of EE. The self-administered Environmental Education Curriculum Needs Assessment Questionnaire (EECNAQ), which was developed by the synthesis of original EE goal levels with their attendant goal statements is adequate for its purpose, the assessment of environmental educators' perceptions.

The need for set criteria or frames of reference to which to relate in assessing curricular needs (and educational needs in general) is obvious, and the researchers established the required frame of reference by using an appropriate (though heavily value-laden) "authoritative" set of goal statements.

The research design, methodology, and procedures employed in the study are in accord with well-established practices and the accepted conceptual framework within a matrix of similar perception assessment studies in the field.

A major issue of concern, however, is that generally speaking, the perceptions of professional environmental educators concerning the state of affairs and "what should be done" in environmental education within on-going science and general education are well known based on the vast amount of available literature in the field (Keiny and Zoller, 1987).

In fact, the goal levels - goal statements used in the study as the frame of reference reflects these perceptions. In this respect, the major findings of the study could have been predicted to begin with.

Regardless of whether or not the major goals of EE which are agreed upon by environmental educators are accepted as vital to the individual and society by the general educational community, society, and social and political policy-makers, the following are some related questions that should be addressed through direct objective research and assessment to serve as a basis for educational decision- and policy-making.

- Are the goals of EE being attained fully or in part by other educational programs and curricula?
- To what extent are EE goals being achieved by existing EE curricula?
- What are the goal levels?

- Do existing EE curricula (and programs) actually pursue the goals of EE in the way these goals are "translated" through them into class practices?
- Based on the outcomes of the above, do we need new EE curricula or should we just improve and modify the implementation of existing ones?
- Can the achievement of long-term objective of EE be demonstrated via practice feasible research?

As a matter of fact, some of the above have been recommended by researchers for further research.

REFERENCES

Hungerford, H. R., R. B. Payton and R. J. Wilke, "Goals for Curriculum Development in Environmental Education." The Journal of Environmental Education, 11 (1980), No. 3: 42-47.

Keiny, S. and U. Zoller, (Eds.), Conceptual Issues in Environmental Education. Heidelberg: Springer Verlag, 1987.