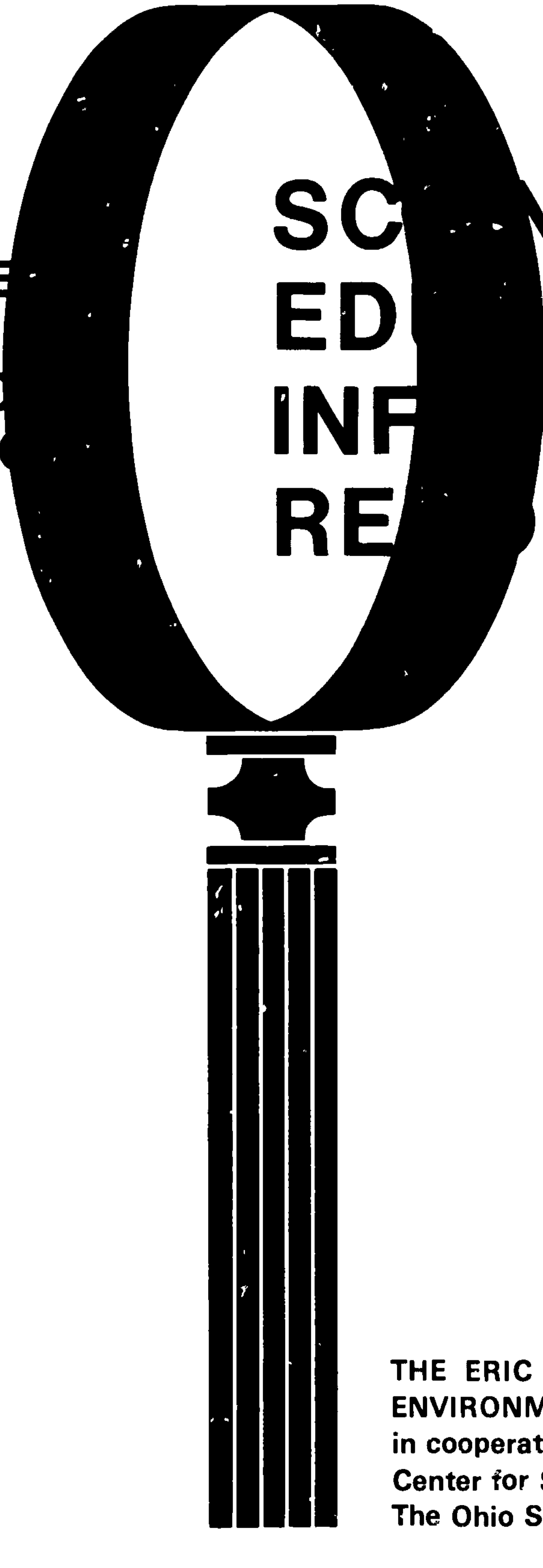
DOCUMENT RESUME

ED 272 393 SE 046 888

| | |
|---|---|
| AUTHOR | Blosser, Patricia E., Ed.; Helgeson, Stanley L., Ed. |
| TITLE | Investigations in Science Education. Volume 12, Number 3. |
| INSTITUTION | Ohio State Univ., Columbus. Center for Science and Mathematics Education. |
| PUB DATE | 86 |
| NOTE | 65p. |
| AVAILABLE FROM | Information Reference Center (SMEAC/IRC), The Ohio State Univ., 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; v12 n3 1986 |
| EDRS PRICE | MF01/PC03 Plus Postage. |
| DESCRIPTORS | Academic Achievement; *Classroom Techniques; College Science; Curiosity; Demonstrations (Educational); *Diagnostic Teaching; Elementary School Science; Elementary Secondary Education; Genetics; Higher Education; *Instructional Improvement; Pacing; *Problem Solving; Science Education; *Science Instruction; Secondary School Science; Student Attitudes; *Teaching Methods |
| IDENTIFIERS | *Science Education Research |

ABSTRACT

Abstracts and abstractors' critiques of six science education research studies and two responses to critiques are presented in this journal issue. Each of the studies addressed some aspect of instruction. Areas investigated include: (1) the effects of teacher-demonstration as compared to those of self-paced instruction on student achievement; (2) use of a teachers' self-report form for the study of classroom transactions; (3) effectiveness of diagnostic-remedial instruction; (4) relationships among elementary school students' interest in science, attitudes toward science, and reactive curiosity; (5) methods employed in solving high school genetics problems; and (6) the effect of instructional intervention to teach sixth grade students experimental problem solving. Responses to critiques on the studies of the teachers' self-report form and experimental problem solving conclude this issue. (ML)



# SCIENCE EDUCATION INFORMATION REPORT



INVESTIGATIONS IN SCIENCE EDUCATION

Volume 12, Number 3, 1986

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

# INVESTIGATIONS IN SCIENCE EDUCATION



National Association for Research in Science Teaching

ERIC® Clearinghouse for Science, Mathematics, and
Environmental Education

INVESTIGATIONS IN
SCIENCE EDUCATION

Volume 12, Number 3, 1986

Published Quarterly by

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

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

Volume 12, Number 3, 1986

NOTES FROM THE EDITOR:

---

Issue number 3 of Volume 12 of Investigations in Science Education contains critiques of six articles in which some aspect of insrruction was investigated and reported. Two responses to critiques are also included; these relate to critiques within this issue.

Eniaiyeju investigated the effects of teacher-demonstration as compared to those of self-paced instruction on student achievement in both concept learning and problem solving. Tamir reported on the use of a self-report form by teachers to obtain information about classroom practices. Yeany and Miller studied the effectiveness of diagnostic-remedial instruction. Harty, Andersen and Enochs looked for relationships among interest in science, attitudes toward science, and reactive curiosity in a study of the affective aspects of instruction. Stewart studied methods high school students used in solving genetics problems. Ross and Maynes investigated the effect of instructional intervention to teach sixth grade pupils experimental problem solving.

In the response section of this issue, Ross has provided some reactions to Za'rour's critique and Tamir has, in response to the critique of his research, included a copy of the self-report instrument he developed.

Patricia E. Blosser
Editor

Stanley L. Helgeson
Associate Editor

INSTRUCTION

Eniaiyeju, Paul A. "The Comparative Effects of Teacher-Demonstration and Self-Paced Instruction on Concept Acquisition and Problem-Solving Skills of College Level Chemistry Students." Journal of Research in Science Teaching, 20 (8): 795-801, 1983.

Descriptors--Chemistry; *College Science; *Concept Formation; Concept Teaching; Conventional Instruction; Demonstrations Educational; Higher Education; Individualized Instruction; *Pacing; *Problem Solving; Science Education; *Science Instruction; Scientific Concepts; *Student Attitudes; Teaching Methods

Expanded abstract and analysis prepared especially for I.S.E. by Carolyn Carter, David V. Frank, and April Gardner, Purdue University.

## Purpose

There appear to be two main purposes of this study. The first is to compare the effects of a self-paced (SP) mode of instruction with a teacher demonstration (TD) mode of instruction on student achievement in both concept learning and problem solving. The second is to compare student preferences for each mode of instruction. Although the author states a third "research question", comparing the effectiveness of the modes of instruction for males and females, he never deals with this issue.

## Rationale

The author uses the terms "individualized" instruction and "self-paced" instruction synonymously. He cites two studies from the 1970's in which self-paced instruction was favored and two studies in which conventional (nonindividualized) instruction was favored. He also cites one other study which noted that the s of studies comparing self-paced with teacher-demonstration methods is generally inconclusive. He believes that these inconclusive results may be due to the different types of courses, cognitive outcomes, and students.

Only a brief description is given for the major types of outcomes the author is trying to measure: concepts and problem-solving. The use of these terms is based entirely on the work of Campbell and Milne (1972). According to the author, Campbell and Milne classify the top three levels of Bloom's taxonomy (analysis, synthesis, and evaluation) as problem solving. Apparently, concepts consist of the lower three levels of the taxonomy.

## Research Design and Procedures

A posttest only control group design (Campbell and Stanley, 1966) was used in this study, as shown below:

R------X(SP)------O(SP)
R------X(TD)------O(TD)

Three instructional units were prepared by the author. These units covered the topics of acids, bases, and salts; solubility; and electrochemistry. The self-paced (SP) units allowed the students to carry out "structured laboratory activities"; these were demonstrated by the teacher in the TD groups. Eight one-hundred minute class sessions were required to complete the instruction.

At the end of the second unit, the SP and the TD groups were switched. The purpose of this rotation was to allow students in each group to experience both methods of instruction. At the end of the three units, students were given a test to measure their attitudes towards the SP and TD methods.

The sample was composed of students who were taking chemistry as a minor subject in the Advanced Teachers' College in Zaria, Nigeria. There were 40 students in the pilot study and 60 students in the main study. The students were split into two groups: those who had secondary school preparation and those who had teacher training education. Students from each group were randomly assigned to the two sections. They were also given the West African Examination objective test in chemistry to demonstrate that the two groups were equivalent in previous chemistry knowledge.

The variables studied were: independent variable--instructional mode (SP vs. TD) and dependent variables--concept learning, problem solving, and preference for instructional mode. A 54-item multiple-choice test containing both concept and problem-solving items was given after the second unit of instruction. The test was validated by a panel consisting of twelve judges plus the author. The classification of each item depended on agreement of 85 percent of the members of the panel. When the test was given to a group of students not in the pilot or main studies, the reliability of the test as measured by the KR-20 was 0.89. The test was given after the second unit of instruction (just before the two groups were switched). A 16-item Likert scale was given at the end of all eight sessions to measure students' attitudes toward the two instructional methods. Neither the reliability nor the validity of this scale was discussed.

Findings

In the pilot study, the SP group mean score on the entire posttest was reported to be significantly higher than the mean score of the TD group (37.51 vs. 30.34). The mean scores from the main study are shown in the table below.

TABLE I

Means and Standard Deviations of the Posttest Results

| Skill | Statistic | Mode of Instruction | |
|---|---|---|---|
| | | SP | TD |
| Concept Learning | M | 13.93[a] | 10.72 |
| | S.D. | 4.0 | 4.42 |
| | N | 60 | 60 |
| Problem Solving | M | 39.50[b] | 31.60 |
| | S.D. | 9.97 | 4.88 |
| | N | 60 | 60 |

[a]Perfect score = 24
[b]Perfect score = 60

The scoring system that was used to derive an 84-point total for a 54-question test was not described. A one-way analysis of variance on the scores in Table One reveals that:

1. Students in the self-paced group significantly outscored the TD group on concept questions ($p < 0.01$).
2. Students in the self-paced group significantly outscored the TD group on problem-solving questions ($p < 0.0010$).

The attitude results from the Likert scale are reported only for the pilot study, although the author states that he used the same procedures to carry out the main study. According to an adjusted chi-square analysis, 81 percent of the 16 items (that is, 13 out of 16 items) received strongly positive responses. The author does not indicate what this statement means, nor does he indicate whether these results are statistically significant.

## Interpretations

The author notes that the self-paced mode of instruction "can profitably be used to teach the areas of chemistry selected for the experiment." Eniaiyeju explains his results in terms of the greater student involvement in the SP group. He compares this with the passive role of the student in the TD group.

He mentions two ions in drawing conclusions from his study: small sample size a n unknown effect due to "teacher competence." The author urges further work to extend or replicate these results.

## ABSTRACTOR'S ANALYSIS

This study was difficult to analyze. This difficulty may be due to the quality of the written report; it could also be due to quality of the execution and interpretation of the experiment. There is evidence from the article that it is a little of both.

The abstract itself provides an immediate example of the inconsistent writing. It implies that the 60 students in the main study were given the attitude test and that most of these students preferred the self-paced instruction. Yet Eniaiyeju actually reports attitude data only for the pilot study, consisting of 40 students. It is not clear from the article if students in the main study were asked to complete the Likert scale; no data were reported. At least the title is more descriptive of the report than the abstract, since it refers only to cognitive objectives.

Another example of the confusing writing style occurs when the author attempts to explain why the teacher-demonstration method "lowered" the students' achievement. No reference point is given for this comparison. Is the TD group being compared with the SP group for the same year? With the TD group of the pilot study? Since the data for the pilot study and the main study are reported in different ways, it is impossible for a reader to deduce what this statement means.

Another confusing point is determining exactly what Eniaiyeju means by "self-paced." It is not clear whether the students in the SP group performed complete experiments, small experimental steps, or just paper-and-pencil exercises.

One of Eniaiyeju's stated purposes is studying the interaction of instructional mode and gender. All of the necessary data appear to be present, yet no reference to gender is made in the results. Why isn't this reported?

Finally, the author's best documented claims are related to the effectiveness of self-paced instruction for teaching concepts and problem solving. Examples of both concept and problem-solving questions are provided. The author's classification of his test questions does not seem to be consistent with the definitions he has adopted.

This test item is characterized as a concept question:

Hard water
(a) will not give a lather with a small amount of soap solution;
(b) is good for irrigation;
(c) contains chlorides, sulphates, and bicarbonates;
(d) good a precipitate on boiling (sic);
(e) is sharp to taste.

This question is clearly a knowledge-level question on Bloom's taxonomy. According to the author's definitions of concepts versus problem solving, this has been classified correctly. However, it is difficult to pinpoint one correct (or even one best) answer to this question.

The author cites the next item as representative of a problem-solving question:

Halima needed 5 g of sodium oxalate for an experiment. She poured some onto a paper and when she weighed it she found that she had poured 15 g. What should she do with the extra salt?
(a) pour the extra oxalate back into the stock bottle;
(b) flush it down the sink;
(c) save it for latter (sic) use;
(d) adjust her experimental need to 15 g;
(e) ask her teacher where to dispose of it.

There would be no problem classifying this question at one of the higher levels of Bloom's taxonomy if it were an essay question, where a student could present a solution to the situation and describe why the solution was appropriate. Since this is not an open-ended question, the abstractors assumed that there could be only one correct answer and tried to determine what that answer was. Only the first response is incorrect; the other four answers seem to be entirely plausible, depending on the teacher's instruction. For instance, if the experiment called for multiple runs, she could save the extra chemical for use in another run. If she had prior knowledge that oxalate was not hazardous, she could flush the extra portion down the sink. Therefore, any instruction which would make it possible for students to choose one correct answer would reduce this question to at most the comprehension level of Bloom's taxonomy. Then the question becomes a concept question according to the author's definition. So the two questions which were given as examples seem to be testing at best the difference between knowledge questions and higher levels, rather than between concepts and problem solving.

Of greater concern than the actual classification of the test items is the basis for that classification. The author's only justification of the definitions of the words "concepts" and "problem solving" comes from Campbell and Milne. Since the time that book was published, these words have developed a much more specific meaning to most science educators. Rather than higher or lower levels of Bloom's taxonomy, a concept has recently been defined as "a regularity in events or objects designated by some label" (Novak and Gowin, 1984). And according to Hayes, "whenever there is a gap between where you are now and where you want to be, and you don't know how to find a way to cross that gap, you have a problem" (Hayes, 1981). While Eniaiyeju need not have chosen these particular definitions, current research does not divide the cognitive domain into simply concepts and problem solving. It was noted earlier that his literature review of these dependent variables was sketchy; perhaps a more thorough review would have led to definitions more in keeping with current research.

A strong point of the study was the experimental design. Subjects were carefully randomized into the two groups. An objective chemistry test administered to all students before the study indicated that the two groups were equivalent in previous chemistry knowledge. A problem that the author mentions about the design is the use of different teachers in the two sections. The SP teacher obviously played some role in the classroom, since the teacher was supposed to act as an adviser and to speed up the work of the slower students. The teacher effects cannot be separated from the effects of the mode of instruction when evaluating the test scores.

In spite of the good design, this experiment does not make a significant contribution to an understanding of individualized instruction. All that one can conclude is that individualized methods were more effective than teacher-directed instruction for three chemistry units in a college setting in Nigeria. Due to the flaws in the construction of the test items, it is difficult to determine whether the SP students had richer conceptual networks and better problem-solving skills or whether they had merely learned more facts. The results are neither generalizable nor informative. Yet the article does suggest some other directions which would have been more useful. Since the study is meant to be developmental as well as experimental, more effort could have been made to describe how self-paced materials were adapted and prepared for a Nigerian setting. This information could be useful for other instructors who are also trying to prepare materials for other settings. Another direction for research is suggested by the literature review, where Eniaiyeju states that individualized instructio may be more effective for certain types of students and courses. However, all that the current study does is add another vote in favor of individualized instruction, without adding any insight into why this mode of instruction is sometimes successful.

## REFERENCES

Campbell, D. T. and J. C. Stanley. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally, 1966.

Campbell, C. V. T. and W. J. Milne. Objective Testing in Chemistry. London: Heinemann Educational Books, 1972.

Hayes, J. R. The Complete Problem Solver. Philadelphia: The Franklin Institute Press, 1981.

Novak, J. D. and D. B. Gowin. Learning How to Learn. Cambridge: Cambridge University Press, 1984.

Tamir, Pinchas. "Teachers' Self Report as an Alternative Strategy for the Study of Classroom Transactions." Journal of Research in Science Teaching, 20 (9): 815-823, 1983.

Descriptors--*Classroom Observation Techniques; Junior High Schools; Science Education; *Science Instruction; *Secondary School Science; *Self Evaluation Individuals; Test Reliability; Test Validity

Expanded abstract and analysis prepared especially for I.S.E. by Constance M. Perry, College of Education, University of Maine at Orono.

## Purpose

The purposes of the study were:

1. to describe the Structured Self Report Form (SLRF);
2. to assess its validity and reliability;
3. to illustrate the kinds of information that can be obtained through the use of the SLRF using the example of science teaching in Israel's secondary schools.

## Rationale

There is a well documented need for information about actual classroom interactions for the purposes of curriculum evaluation and research on teaching. Direct observation and questionnaires administered to teachers and/or students are the two major approaches to the study of classroom transactions. Although both approaches can provide useful results, they suffer from serious limitations. Direct observations are very expensive and time consuming and the kind of information obtained depends on the bias of the instruments and/or of the observer's interpretations. Questionnaires are relatively inexpensive but tend to provide general and idealistic views. This study proposes an alternative. An instrument entitled the Structured Lesson Report Form (SLRF) was designed. Teachers provide a detailed report on a specific lesson. It is assumed that, as one collects a sufficient number of SLRF's, valid and reliable information about classroom practices can be obtained. The study was designed to test that assumption.

## Research Design and Procedures

### Method and Sample

The Structured Lesson Report Form (SLRF) is completed by the teacher following the lesson. Most items are structured so that the teacher must select the appropriate option, however, certain items are open but carry clear instructions. For example, "What is the kind of homework assigned?" Information is obtained on lesson topic, description of students, classroom setting, inquiry features, teacher's satisfaction, etc. The SLRD also requires the teacher to describe in chronological order what happened in the class.

Two hundred and fifty SLRF's were obtained from 250 teachers. The lessons were:

200 junior high - 102 biology
48 physics
36 chemistry
50 senior high - all biology

The teachers were 70% female, 21% had less than two years teaching experience, 50% had six or more years teaching experience. Sixty percent held a university degree and the rest had been trained at teacher education colleges.

In addition, 90 SLRF's were collected by 20 external observers. A sample of 40 lessons that were reported by both teachers and observers was used to establish reliability.

Data Analysis. Structured items were analyzed for frequencies. Responses to open items were categorized and frequency counts were made for each category. In addition, three notices were made for each category:

(1) whether it occurred at the beginning, middle or end of lesson;
(2) whether it occurred once or more during the lesson; and
(3) whether it was made by the teacher, student(s) or by both.

Frequency counts were made for each category over each of the three descriptors.

## Findings

Validity. Five science educators found the categories of SLRF to be adequate descriptors of classroom transactions.

Reliability. The results of the 40 SLRF's completed by both teachers and observers were analyzed and comparisons made for each category. High levels of agreement were reached in certain categories (94 percent for use of audio-visual aids, 86 percent for nature of homework, 78 percent for who is doing what and when in the lesson). Medium levels of agreement were reached in categories which required a high level of inference such as the occurrence of learning by inquiry (60 percent) or assessment of students' satisfaction and enjoyment (56 percent). The average agreement between observers and teachers was 80 percent.

Classroom Environment. About 60 percent of the lessons took place in a laboratory while the rest were held in regular classrooms. At the junior high level, 44 percent were recitation, 28 percent labs and 28 percent mixed. The junior high classes were classified as 15 percent low ability, 55 percent mixed ability, 30 percent high ability. The distribution at the senior high level was 6 percent low, 22 percent mixed and 72 percent high ability.

Eighty-five percent of the laboratory lessons took place in labs. In labs students worked in pairs (50 percent senior high, 36 percent junior high) and in the rest of the lab lessons students worked in teams of three to four.

Instructional Resources. Thirty-eight percent of junior high lessons were planned with the aid of the teacher's guide and 65 percent followed the sequence recommended by the teacher's guide. In physics and chemistry, two books were used by more than 60 percent of the classes, yet the most widely used textbook in biology was used by fewer than 30 percent of the classes. The most widely used instructional aid was the chalkboard followed by charts and diagrams.

More assistance of laboratory technicians was used in biology classes and at the senior high level, in a third of the lessons the technicians actually assisted in instruction. Most labs were taken from the textbooks.

In more than 60 percent of the science classes inquiry was moderately or highly experienced. Homework was routine with physics and chemistry teachers relying more heavily than biology teachers on textbook questions. Reading assignments were much less frequent than writing assignments.

Classroom Transactions. Forty-two percent of the lessons were integrated recitation and lab, 23 percent pure lab and 35 percent pure recitation.

The predominant components of a typical lesson were:

(1) posing a problem by the teacher
(2) explaining, more than half of it done by the teacher but with more collaboration with students in recitation
(3) discussing
(4) writing in notebooks.

In laboratory and integrated lessons students commonly designed experiments with teacher help, performed them with other students and pooled results with the guidance of the teacher.

Homework assignments were common but rarely utilized or built upon. A relatively small proportion of class time was devoted to reading from the text. In only a small percentage of the classes were students assigned text reading as homework. Referring to related literature or reading relevent literature as homework was quite limited.

## Interpretations

A structured format was used to obtain descriptive information and an open ancedotal format for the report on actual classroom transactions. The analysis of the anecdotal portion was tedious but since it saved not having to conduct direct observations the extra effort was worthwhile.

The comparison of teachers' own responses with that of observers suggests that, at least with regard to the SLRF categories, teachers' reports are quite reliable. The validity and reliability of the data are significantly enhanced because teachers are asked to relate to one lesson rather than make generalizations about many lessons. This approach is feasible and worthwhile especially when circumstances preclude the use of more expensive methods. SLRF may be used with a larger sample to complement and verify data obtained by direct observation of a smaller sample.

The data suggests that, unlike the U.S., science classes in Israel are, on the average, inquiry oriented, have an abundance of discussions, explanations and hands-on laboratory activities.

## ABSTRACTOR'S ANALYSIS

The topic of the Tamir article, a teacher self report device for the study of classroom transactions, is timely. The study of classroom transactions, although certainly not new, has been especially evident in the last several years. The recent research in areas including classroom interaction, time-on-task, teacher expectations, questioning and management are all examples of studies of classroom transactions. Most of the studies in the areas just mentioned employed direct observation, a method that as Tamir says, is expensive and open to bias of the instrument and/or observer. His study to develop a less expensive, reliable, and valid means to study classroom transaction is pertinent and valuable; however, the report of the study does raise several questions and could benefit from clarification.

The purposes of the study were:

(1) to describe the Structured Self Report Form (SLRF);

(2) to assess the validity and reliability;

(3) to illustrate the kinds of information that can be obtained through the use of SLRF using the example of science teaching in Israel's secondary schools.

The first purpose could have been better met by including a SLRF either in the text or as an appendix. Although the description gives examples of questions and the tables of results later on help fill in the readers' knowledge of the SLRF, a complete description at the beginning would be helpful.

The description provided raised other questions as well. Tamir states that information obtained when employing direct observations... "greatly depends on the bias of the instruments used..." Certainly instruments are biased as soon as they ask for any kind of information and do not collect every possible piece of information. The same must be true of the SLRF, as it too asks for specific data. Even the open-ended part of the SLRF which calls for a complete description of what happened in the lesson is subject to the bias of the observer just as Tamir says direct observations are. Classrooms are extremely complex.

> In a single day, a teacher may engage in more than 1000 interpersonal exchanges with students (Jackson, 1968). Teachers in secondary schools may have to remember the interactions they had with over 150 students a day. Not only do teachers have many interactions with students, they also have to interpret complex classroom behavior on the spot (Levine and Mann, 1981; Brophy et al., 1981; Anderson-Levitt, in press) (Good and Brophy, 1984, pg. 21).

A teacher cannot remember, even after one lesson, everything that happened; therefore the description has to be biased.

A more practical question is how long does it take a teacher to complete a SLRF? Although it appears reasonable and valuable to have many teachers fill out one SLRF and thereby obtain an overview of classroom transactions, could the SLRF realistically be used several times by individual teachers or would the teachers be unable to appropriate the time regularly?

The validity and reliability statements require clarification. The article states that five science educators who examined the categories of SLRF found them to be adequate descriptors of classroom transactions. What did the science educators judge the categories against? Against what criteria or models?

The description of the reliability process is confusing. SLRF's for the same forty lessons were completed by both teachers and observers. Did the observers complete the SLRF's during the lesson or after the lesson? The establishment of reliability would be stronger if observers completing SLRF's during a lesson highly agreed with teachers completing SLRF's after lessons. If both groups completed the forms after the lessons, both groups could be equally guilty of generalizing a description. The article does not state which was the case.

Only some of the reliability scores (levels of agreement) are noted. It would have been meaningful to see them for each category. The average level of agreement (80 percent) is helpful but gives us no idea of the range. Is the 56 percent agreement listed for the category of students' satisfaction and enjoyment the lowest level or not?

The third stated purpose of the study (to illustrate the kinds of information that can be obtained through the use of the SLRF) comprised the majority of the article. It might have been more appropriate to more fully explain the procedures and results in regard to purposes one and two and shorten the illustrations of kinds of information that can be obtained, or separate purpose three into another article.

The results as described for purpose three are somewhat difficult to follow. There is much overlap among the three areas: classroom environment; instructional resources; and classroom transactions. For example: percentages of the type of lessons (lab, recitation, integrated) are provided under both the classroom environment and transaction categories. They are different percentages. Under classroom environment percentages are given for junior high recitation, lab and integrated lessons but none are given for high school. In the transactions section percentages for each lesson type are stated but it is not stated whether the percentages are for the junior high, senior high or total. It is confusing. Under instructional resources it is stated that 38 percent of the lessons in the junior high were planned with the aid of the teacher's guide and 65 percent followed the sequence recommended by the teacher's guide. It is assumed that the 38 percent is part of the 65 percent (since 38 percent and 65 percent total 103 percent) rather than in addition to, but what does that mean? Could 65 percent follow the sequence recommended by the teacher's guide but only 38 percent actually be planned using the guide?

The article stressed the high level of inquiry in the science classrooms. The inquiry category of the SLRF had three choices: hardly existing, moderately experienced and widely experienced. The results were impressive but could have been more so if the term inquiry was defined to include its parameters and descriptions were supplied for each of the three choices.

The researcher concludes that the use of the SLRF to study classroom interactions is feasible and worthwhile especially when circumstances preclude the use of more expensive methods (direct observations). A suggestion is also made that the SLRF might be used with a larger sample to complement and verify data obtained by direct observations. This practice would provide for improved generalizability. The conclusion and suggestion appear appropriate. The SLRF certainly is usable for collecting data on classroom transactions and allows for summaries of such to be made. As with any report form it is more appropriate for certain areas of research. The SLRF allows the use of large samples to obtain an overview of practices but would not be appropriate for studies of precise teacher or student behaviors such as number and type of questions asked and to whom, to state just one example. Keeping that limitation in mind and the comments made in regard to the validity and reliability, the SLRF still appears to add another worthwhile dimension to the study of classroom transactions.

## REFERENCE

Good, T. L. and J. E. Brophy. Looking in Classrooms, 3rd Edition. New York: Harper and Row, Publishers, 1984.

Yeany, Russell H. and P. Ann Miller. "Effects of Diagnostic/Remedial Instruction on Science Learning: A Meta Analysis." Journal of Research in Science Teaching, 20 (1): 19-26, 1983.

Descriptors--*Academic Achievement; *Diagnostic Teaching; Elementary School Science; Elementary Secondary Education; Feedback; *Remedial Instruction; *Science Education; *Science Instruction; Secondary School Science; *Teaching Methods

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

## Purpose

The purpose of this study was to integrate the existing science education research relating to the effectiveness of diagnostic/remedial instruction to:

a) determine the magnitude of the effect that can be expected from diagnostic/prescriptive instruction:

b) determine the effects attributable to different components of the method;

c) examine the relationship of the effect and school level; and

d) examine the relationship between the effect and the number of times the students experienced a summative test within the study period.

## Rationale

Bloom (1968) emphasized the importance of measuring learners' progress to diagnose difficulties and providing the learners with feedback and prescriptive remediation. Okey (1974) identified the components of a diagnostic/remedial instructional system as:

Step 1: Specify performance objectices

Step 2: Develop diagnostic measures for objectives

Step 3: Teach using any preferred procedures

Step 4: Test achievement of objectives using diagnostic measures

Step 5: Remediate and rediagnose, if desired

Step 6: Administer summative test

Numerous experimental studies of diagnostic/prescriptive science instruction have been reported in the last fifteen years. These studies have used a variety of treatments and dependent variables. Some studies provided diagnosis and feedback but no remediation. Others used both diagnosis and remediation. Still others compared several treatments: no diagnosis, diagnosis only, and diagnosis with remediation. Usually the dependent variable was the learning of some science fact, concept, or skill. Attitudinal measures were of secondary importance or ignored altogether.

In most cases the results of these studies favored some type of diagnostic feedback, prescribed remediation, or both. However, some studies did not find a significant effect, while still others found significant differences for some units of instruction and not for others.

The present study was conducted to attempt to synthesize the results of these studies to determine the impact of diagnostic/prescriptive instructional strategies on the teaching of science.

## Research Design and Procedure

The research literature was scanned using computer and manual search techniques to identify studies whose titles or descriptors suggested that they pertained to diagnostic/remedial instruction in science. A total of 28 studies were located. These documents were studied to ensure that each study to be included in the meta-analysis met the following criteria:

a) treatment had to include at least the diagnostic testing component of the diagnostic/remedial instructional strategy;

b) the content focus of the instruction had to be science achievement or attitudes; and,

c) enough statistical data had to be given or be available from the author to derive effect sizes for the treatment examined in the study.

The diagnostic/remedial treatments used in these studies were classified into three categories:

| Type | Definition | # of Studies |
| --- | --- | --- |
| I | no diagnosis or remediation | 20 studies |
| II | achievement diagnosis with feedback only | 9 studies |
| III | both diagnosis and prescribed remediation | 18 studies |

Diagnosis was defined as any assessment of a student's progress related to a specific of instructional goals that occurred before the summative measure of those goals. Prescribed remediation included any instructional assistance or guidance the student received that was designed to strengthen learning weaknesses revealed by the diagnosis. Remediation included such things as: assignment of remedial materials or activities and individual assistance. They could be self, peer, or teacher administered.

The impact of the various treatments was assessed using the "effect size, ES" (the difference between the means of the control and treatment groups divided by the standard deviation of the control group). Where there was not sufficient information provided in the study to calculate the effect size, the author was contacted. Where more than one dependent variable was measured, a separate effect size was calculated for each. Thus the unit of analysis was the finding and not the study.

Mean effect sizes for the Type II and III treatments were calculated by comparing the treatment and control groups within a study and then averaging across all the effect sizes to a particular class of treatment. To assess the relationship between diagnostic/remedial instruction and school level and frequency of summative testing, an $R^2$ for each of the possible relationships was calculated.

## Findings

The results of the meta-analysis on the effectiveness of the two treatment types are summarized below:

| | Type II Feedback Only | Type III With Remediation |
|---|---|---|
| Mean ES | 0.49 +/- 0.12 | 0.55 +/- .07 |
| Minimum ES | -0.11 | -0.25 |
| Maximum ES | +1.99 | +1.93 |
| n (findings) | 19 | 49 |
| n (+ findings) | 16 (84%) | 45 (92%) |
| n (- findings) | 3 (16%) | 4 (8%) |

The mean effect size for both types of treatment were significantly different from zero ($p < 0.05$) and the number and size of the positive findings indicates that the treatments had a beneficial effect on student achievement. However, the effects of the two treatment types were not significantly different from each other.

Attitudinal measures were only available from studies that investigated Type III treatments. No significant effects on attitudes towards instruction or science were found.

No significant effects were found when findings were compared on the basis of school level (college and precollege), or the length of time the students had been exposed to diagnostic/remedial instruction.

## Interpretations

On the basis of their findings the authors concluded that achievement can be significantly and positively influenced by diagnostic/remedial instruction. The magnitude of this effect can be expected to be about 0.55 standard deviation units of achievement when compared to an instructional strategy not employing diagnosis and remediation.

Since there was no significant difference between the effects of diagnostic feedback only (Type II) and diagnosis and prescription (Type III), the authors conclude that science students are capable, in the absence of prescribed remedial activities, of attending to their own remediation when provided with feedback from the diagnosis of achievement deficits.

The authors feel that these results are encouraging because it is much more practical for classroom teachers to provide their students with diagnostic feedback than to provide complex remediation schedules and cycles.

## ABSTRACTOR'S ANALYSIS

Over the years integrative reviews of educational research have served the educational community well. They have provided a convenient summary of a large body of research for educational practitioners and others wanting to get an overview of the current "state-of-the-art" in a particular research area. In recent years these reviews have become more quantitative due to the application of the statistical techniques for meta-analysis described by Glass (1978). These more quantitative techniques have seemed to appeal to science educators because the recent literature in science education contains many reviews of this type.

The current interest within the educational community in competency-based education, mastery learning, and developing adaptive, microcomputer-based instructional systems means that the Yeany and Miller's choice of diagnostic/prescriptive instructional systems for review was a timely and appropriate one.

The validity and utility of any review of the research literature is a function of the criteria used to include or exclude studies and the thoroughness of the search of the literature. Yeany and Miller chose to restrict their review to studies that pertained to diagnostic/remedial instruction in science. This way of restricting their analysis seems peculiar because learning theorists, although they may disagree about the particulars, all classify student learning into categories such as paired associates, concepts and rules rather than by subject area such as science or mathematics. The results of the review would be more generalizable if the authors had not restricted their study in this seemingly arbitrary way.

Another factor affecting the validity of a review of the literature is the thoroughness of the search of the literature. The authors state that they used computer and manual search techniques to locate documents pertaining to diagnostic/remedial instruction in science but they neglect to mention which sources and data bases were scanned. Since other contemporary reviews (e.g., Lysankowski and Walberg, 1982) do provide this information, readers have come to expect it. Without it a reader is unable to assess the thoroughness of the literature search.

Although the authors do not mention the specific search terms or descriptors they used in conducting their search of the literature, they appear to have focused almost exclusively on diagnostic/remedial instruction. This term is frequently used in special education. In this context it is used to refer to the diagnosis of learning disabilities and the prescription of instructional treatments to eliminate or circumvent them (Arter and Jenkins, 1979). As a result, many researchers interested in mastery learning and feedback avoid the use of the term diagnostic/remedial instruction in reference to their work. The use of broader search terms would have uncovered a larger number of studies and review articles such as Barringer and Gholson (1979). This is particularly unfortunate in view of their finding regarding the importance of corrective feedback.

In general, the report was clearly and succinctly written. The tables and illustrations were neat and easily interpreted. However, their conclusion that "it appears that achievement can be significantly and positively influenced through diagnostic/remedial instruction" appears to be somewhat sweeping in its generality in view of the fact that the authors restricted their reviews to studies relating to the learning of science.

Despite the study's limitation, its findings are consistent with Lysakowski and Walberg's (1982) meta-analysis. Both reviews indicate that providing corrective feedback enhances student achievement. This is an important finding deserving of widespread dissemination because of its practicability in normal classroom situations. Researchers should be encouraged to explore this area to determine specifically which types of learning outcomes (facts, concepts, etc.) are most favorably affected by corrective feedback and which types of feedback are most effective.

The statistical analysis of the studies that were included in the review appears to have been competently done in accordance with normally accepted procedures. However, when reading through this report and the one by Lysakowski and Walberg (1982), this reviewer felt a pang of nostalgia for the "old style" review in which the reviewer carefully pointed out the strengths and weaknesses of the various studies being considered. The more modern meta-analytical technique involves summing and averaging the effect sizes of the different studies and, as a result, each study's effect size contributes equally. Although an individual study's effect size is indirectly influenced by such things as sample size and the reliability of the testing instruments used, it does not reflect such important factors as the strength of the research design, the validity of the tests used, or the length of the treatment. We may be losing more than we are gaining by relying heavily on strictly quantitative analyses. Perhaps we should attempt to write research reviews that combine the best aspects of the new and old.

## REFERENCES

Arter, J. A. and J. R. Jenkins. "Differential Diagnosis-Prescriptive Teaching: A Critical Appraisal." Review of Educational Research, 49 (3): 517-555, Fall, 1979.

Barringer, C. and B. Gholson. "Effects of Type and Combination of Feedback upon Conceptual Learning by Children: Implications for Research in Academic Learning." Review of Educational Research, 49 (3): 459-478, Summer, 1979.

Glass, G. V. "Integrating Findings: The Meta-analysis of Research." in Shulman, L. (Ed.), Review of Research in Education, Itasca, IL: Peacock Publishing, 1978.

Lysakowski, R. S. and H. J. Walberg. "Instructional Effects of Cues, Participation, and Corrective Feedback." American Educational Research Journal, 19 (4): 559-578, 1982.

Okey, J. R. "Altering Teacher and Pupil Behavior with Mastery Teaching." School Science and Mathematics, 74 (6): 530-535, 1974.

Harty, Harold, Hans O. Andersen and Larry G. Enochs. "Exploring Relationships Among Elementary School Students' Interest in Science, Attitudes Toward Science, and Reactive Curiosity." School Science and Mathematics, 84 (4): 308-315, 1984.

Descriptors--*Curiosity; Elementary Education; *Elementary School Science; Grade 5; Science Education; Science Instruction; *Science Interests; *Student Attitudes

Expanded abstract and analysis prepared especially for I.S.E. by Ann E. Haley-Oliphant, University of Cincinnati.

## Purpose

The authors' primary purpose of this study was "to determine whether statistical relationships exist among three conceptually related affective variables" (p. 308). The variables studied included: interest in science, attitudes toward science and reactive curiosity.

## Rationale

The rationale for this study was based on a literature review which revealed a necessity for investigating the relationships among student affective variables of interest, attitude and curiosity. Specifically, the research indicated that one of the goals of primary and secondary teachers has been the "nurturing of pupil interest in science, attitudes toward science, and curiosity" (p. 309).

The work done by Kerlinger (1968) on the "directive state theory" of attitudes; Berlyne's (1954) "epistemic curiosity" theoretical configuration and Witty's (1961) conceptual framework and theoretical writings were used to support this study. In this study, attitudes were defined as "certain regularities of an individual's feelings, thoughts and predispositions to act towards some aspect of the environment" (Secord and Backman, 1964, p. 97); reactive curiosity was defined as: "(1) a tendency to approach and explore relatively new stimulus situations; (2) a tendency to approach and explore incongruous, complex stimuli, and (3) a tend. icy to vary stimulation in the presence of frequently experienced stimulation" (Penney and McCann, 1964, p. 323); and interests were defined as "patte.'ns of choice among alternative patterns that demonstrate some stability over time and do not appear to result from external pressures" (Rust, 1977, p. 132).

## Research Design and Procedure

The th.ee dependent variables were interests, attitudes and reactive curiosity. These were used in correlational methods and analyzed by way of the Pearson product moment technique. The dependent variables were measured using the instruments found in Table 1. Each of these modified instruments were direct descendants from earlier instruments. In all cases, the instrument's readability was reduced to the elementary reading level. The authors field tested the instruments with 171 fifth grade children to establish internal consistency reliability. The researchers subsequently administered the three instruments to the study sample of 91 fifth grade students, 44 girls and 47 boys.

TABLE 1

Instruments Used To Measure the Dependent Variables

| DEPENDENT VARIABLE | INSTRUMENT UTILIZED | CHARACTERISTICS OF INSTRUMENT | RANGE OF ITEMS/SCALE |
|---|---|---|---|
| Interest in science | "Children's Interest in Science Measurement" (Meyer, 1969) | 20 Likert-type items | 5 (fully agree) to 1 (do not agree) |
| Attitudes toward science | "Children's Attitude Toward Science Survey" (Fisher, 1973) | 20 Likert-type items | 5 (strongly agree) 1 (strongly disagree) |
| Curiosity | "Children's Reactive Curiosity Scale" (Penney, 1964) | 40 true-false items | |

## Findings

The following significant positive correlations were found:

TABLE 2

| CORRELATION | CORRELATION COEFFICIENT | | |
|---|---|---|---|
| | OVERALL | GIRLS | BOYS |
| Interest in Science-Attitude Toward Science | 0.58 $(p < 0.001)$ | 0.66 $(p < .001)$ | 0.48 $(p < .007)$ |
| Interest in Science-Curiosity | 0.47 $(p < .001)$ | 0.52 $(p < .004)$ | 0.41 $(p < .02)$ |
| Curiosity-Attitude | 0.40 $(p < .002)$ | 0.42 $(p < .02)$ | 0.38 $(p < .03)$ |

In analyzing the results and the instruments, the researchers found the following:

1. The items on the interest measure dealt with either science or scientific application. The items on the curiosity and attitude measure focused on non-science related items as well as a broader array of topics. The researchers state: "It is possible that the attitudes toward science and curiosity instruments might have measured these factors and something else" (p. 313).

2. There were high average scores for personal activity-oriented items, such as "using a telescope." These personal activity-oriented items also got higher ratings if the students were directly involved in the activity as opposed to passive participation, such as "reading a science book; watching a science television show."
3. Both high- and low-interest students in science would favor more active involvement in science.

Interpretations

1. "The statistical relationships among the three variables for both sexes and all students must still remain inconclusive and open for further investigation" (p. 311).

2. The researchers suggest that a much larger sample be used to determine if a strong relationship might exist between these three affective variables.

3. The development and validation of a specific scientific curiosity measure for use by science education researchers and teachers is definitely needed.

4. Further areas of research include:
   a. correlational studies between each or all of these variables and achievement in science, self-concept in learning science, locus of control, intolerance for ambiguity, etc.
   b. ethnographic studies to observe and interview students to shed light on the relationship among these variables.
   c. experimental-control groups, discriminant analysis situations examining cause-effect relationships dealing with these variables.
   d. determine if these three variables could serve as non-traditional identification procedures for science-gifted students.

ABSTRACTOR'S ANALYSIS

1. Relationship of study to other studies.

The literature review comprises several landmark studies in the area of affective research; however, only 10 percent of the literature review is from 1980 to present. Although the area of affective research is in its infancy, the literature review could have been strengthened by citing newer studies that have addressed science attitudes, interest and curiosity.

2. New conceptual contributions.

This study served to confirm "hunches" that many science educators have held over the years even though the findings of this study are inconclusive as to a strong relationship existing among the affective variables. The strength of this study lies i- the suggestions offered for future research in the area of affective studies.

3. Methodological contributions of the study.

No new procedures were utilized. The instruments used and the data analysis are quite common. Interviews with some of the children and teachers might have provided further insight into the correlational relationship; however, this would have extended the scope of the survey beyond the original intentions of the authors.

4. Validity of the study.

There is little if any information provided regarding specific procedures, instrument administration and data collection. This study would be difficult to replicate by other researchers. The following questions can be raised regarding the experimental design of this study:

a. How was the sample selected?

b. Why were fifth graders selected as the subjects of the study?

c. How do we know that the fifth graders selected were a representative sample of the entire population?

d. How were the field test participants selected?

e. What are the specific demographics of the students? If this study was done in an economically disadvantaged area, how might the data be affected?

5. Comments on research design and adequacy of the written report.

The design that is described seems appropriate for the stated objectives. Even though the three instruments are available from other sources, it would have strengthened the written report if more items of each instrument had been provided. The strongest aspect of this study was the extensive listing of the future research possibilities.

Some of the data that were collected were not represented adequately. For example, the researchers state that "items which students rated higher were items that would involve them more actively" (p. 312). Even though this was meant to be a correlational study, the written report would have been strengthened by giving specific numerical data to support this statement.

The researchers briefly address the "Third Assessment in Science of the National Assessement of Educational Progress" (NAEP) data that revealed a decline in attitudes and interests in science. This correlational, exploratory study could have been enhanced by drawing data from NAEP and relating it to the findings of this study. For example, how similar were the items on the instruments to the NAEP instruments? Were similar trends found in both studies regarding the presence of a relationship between the affective variables? Do the findings of this study further substantiate the findings of NAEP?

6. Assessment of the current state of research in the area of study.

The researchers address the need to determine if a statistical relationship exists among the affective variables. Evaluating the affective component in science education is an ongoing problem due to the lack of a clear definition of the three variables. Even though the authors present three definitions of the variables, it is easy to see parallels between and among them. These three variables are typically goals of science education, yet they are somewhat nebulous to measure. Also, is it really possible to measure curiosity without interest or attitude factoring into the instrument? The importance surrounding these affective variables is not really the differences between them but what causes children to develop interest, curiosity and positive attitudes towards science. In addition, it seems logical that these affective variables may have an impact on success in science, selection of science as a career and the perception of science as a useful, necessary body of knowledge that facilitates the everyday living and decision-making. The researchers point to the need for future studies in these areas.

REFERENCES

Berlyn, D. E. "A Theory of Human Curiosity." British Journal of Psychology, 45: 180-191, 1954.

Fisher, T. H. "The Development of an Attitude Survey for Junior High School Science." School Science and Mathematics, 73: 647-652, 1973.

Kerlinger, F. N. "Progressivism and Traditionalism: Basic Factors of Educational Attitudes." Journal of Social Psychology, 48: 111-135, 1958.

Meyer, G. R. A Test of Interests. Milton (Queensland) Australia: Jacaranda Press, PTY LTD, 1969.

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

Rust, L. W. "Interests." Motivation in Education. S. Bell (ed.). New York: Academic Press, 1977.

Secord, P. F. and C. W. Backman. Social Psychology. New York: McGraw-Hill, 1964.

Witty, P. "A Study of Children's Interests." Education, 82: 39-45, 1961.

Stewart, James. "Student Problem Solving in High School Genetics." Science Education, 67 (4): 523-540, 1983.

Descriptors--*Biology; *Genetics; High Schools; *Problem Solving; Science Education; *Science Instruction; *Secondary School Science

Expanded abstract and analysis prepared especially for I.S.E. by A. Line, H. J. Smith, and C. L. Mason, Purdue University.

## Purpose

This study was instituted to investigate the specific steps (procedural knowledge) that high school biology students use when solving monohybrid and dihybrid cross problems, and to describe the extent to which students could justify their execution of each step in terms of their conceptual knowledge of genetics and meiosis.

## Rationale

Underlying this research is the assumption that problem solutions which students can justify are a desired outcome of science instruction. Instruction and evaluation could be influenced by having available descriptions of students' problem-solving difficulties and their misconceptions.

## Research Design and Procedure

Two groups of high school biology students from a single school participated in the study. The first group contained 12 students enrolled in a freshman biology course; the second group contained 15 students who had completed the freshman course and had elected to take an additional general biology course rather than physics, chemistry, or advanced biology for their second science credit.

The data on student procedural and conceptual knowledge were gathered in two audio-taped interviews. The interviews were a combination of the thinking-aloud procedures and the clinical interview format. A combination of these two qualitative techniques was used because the students had previously been exposed to the types of problems used in the interviews and could hence use algorithms to solve them.

In the first interview students were asked to solve a monohybrid problem, a dihybrid problem, and a trihybrid gamete generation problem. In the second interview students were asked to demonstrate the process of meiosis beginning with a parental cell containing three chromosome pairs. For each of the problems, students were asked to think aloud while working the solution. After each problem was solved, the students were asked questions about the meaning of concepts such as gene, allele, heterozygous, homozygous, gamete, recessive, and dominant; the way that they understood these concepts to be related; and, most importantly, to what extent they could justify the execution of their problem steps in terms of knowledge of a mechanism, meiosis. Heavy emphasis was placed on this process during the interview.

The interviews and data analysis were based on five subgoals which the researcher considered essential to correct solution of the problems:

1. construction of a symbolic key to alleles
2. determination of parental genotypes
3. construction of gamete types
4. determination of offspring genotypes
5. determination of offspring phenotypes.

Transcription of the audio tapes from the interview, and the notes that students made as they solved the problems formed the basis for the data analysis. Students' procedural knowledge was represented in a format which could be compared to the procedural representation.

## Findings

After the analysis of individual cases, students were grouped according to the similarities in procedural and conceptual knowledge.

Group A: three students who did not obtain correct solutions to either the monohybrid or the dihybrid problem. These students failed to execute one or more procedural steps. Their procedural difficulties were tied to poor conceptual knowledge. They were unable to assign functional symbols to the alleles and misunderstood basic genetics concepts such as heterozygous, homozygous, dominance, and recessiveness.

Group B: seven students who obtained correct solutions to the monohybrid problem but not to the dihybrid problem. Again, conceptual misunderstandings were at the root of the procedural errors. Their poor conceptual knowledge was particularly apparent with respect to meiosis and the independent assortment of non-homologous chromosome pairs.

The greatest difficulty students in groups A and B had was justifying what they had done when generating gametes in terms of the mechanism of meiosis. For example, when asked (1) how many chromosomes were involved in the dihybrid cross just completed, and (2) how the letters being used related to chromosomes, a number of students responded that they could not tell.

Group C: seventeen students who obtained correct answers to both the monohybrid and the dihybrid problems. These students executed problem-solving steps that were, in nearly all cases, identical to those in the procedural representation. Students in this group were better able to explain and justify their procedural steps than were students in groups A and B. However, only three of 17 students demonstrated understanding by invoking meiosis as a mechanism.

## Interpretations

The most obvious and not unexpected finding was the existence of two groups of students: those who could solve dihybrid problems and those who could not. However, the large number of students who obtained the correct answer but who demonstrated a very limited understanding of the conceptual knowledge of genetics and meiosis is most disturbing. It shows how easy it is to be successful, since success is usually measured by correct answers, by following algorithms without understanding the underlying concepts. More emphasis should be placed on the meanings of central concepts instead of vocabulary memorization. Furthermore, greater attention should be given to explaining to students the relationships among the concepts of a course, unit, or lesson.

Further study on the influence of student evaluation on meaningful learning is warranted. Would meaningful learning increase if what was rewarded was not merely the application of an algorithm? Descriptions of missing concepts and relationships such as this study reveals can be used to help teachers design more effective instruction.

## ABSTRACTOR'S ANALYSIS

Reports on research into problem-solving have been published regularly, and with increasing frequency during the last five years, particularly in the math education field. The Journal for Research in the Teaching of Mathematics has a section devoted entirely to problem-solving. Interest in this area was intensified in 1980 when, at the annual conference, the National Council for Teachers of Mathematics stated that problem-solving was one of the areas in which students are particularly weak and that research in this area was badly needed.

A number of studies have previously been reported in math, physics, chemistry, and chemical engineering. However, the research in these areas defines a "problem" as something unfamiliar to students, for which they have not learned an algorithm. It appears, though this is not explicitly stated, that the students in this study were taught the "ideal procedural solution" outlined by the researcher. The title is somewhat misleading, since the author's use of the term "problem-solving" is inconsistent with the literature.

A significant finding of this study is the lack of integration of different concepts in a subject area (in this case, crosses and meiosis) by students. This was perhaps foreshadowed by the fact that the subgoals (which may have been an instructional template) did not include an explanation of the relationship between genetics and meiosis. Lack of integration is a finding common to many of the studies in problem-solving and appears to be one of the areas in which teachers can help by explicitly discussing the connections with students.

This study is purely descriptive. Responses of a small number of students in a single school were collected and collated. The results cannot, therefore, be generalized, although it seems likely that similar results would be obtained from a larger sample of subjects from different schools in different areas. (The data tables contain two confusing typographic errors: in Table II subgoal D, the fourth gamete of parent 1 is incorrect; and the Table VI footnote should refer to Table V, not Table IV.)

The thinking-aloud technique and clinical interviews are effective methods of gaining information about students' problem-solving processes. Paper and pencil tests often produce only an answer, providing little informaiton about the procedural steps, and no information about the students' understanding of conceptual knowledge. In this study, the extent of questioning during an interview was an individual decision made during the interview. Guidelines for the use of the interview questions would enable more consistent duplication of student interviews and thus enable valid cross-group comparisons in future studies.

More information about the subjects' backgrounds would facilitate extension of this study. Were they taught by the same teacher? What fraction of students in each group were in their freshman year? How were they selected? Were they typical students? Were both freshmen and sophomores studying genetics at the time the study was performed? (At least one student was.) What kind of instruction (content and method) did the students receive in genetics and meiosis, and how closely were the two taught? Answers to these questions would allow better interpretation of the results and more information regarding the development of relationships between different concepts in the subject area.

The techniques used in this research were effective in detecting students' disturbing lack of conceptual integration of meiosis and genetics. Further studies in this area could investigate the ability of different teaching methods necessary for increasing students' development of relationships between concepts, and thus point to the most effective teaching method(s) for this content area.

## REFERENCES

Gabel, D. L. and R. D. Sherwood. "Facilitating Problem Solving in High School Chemistry." Journal of Research in Science Teaching, 20: 163-177, 1983.

Larkin, J., J. McDermott, D. P. Simon, and H. A. Simon. "Expert and Novice Performance in Solving Physics Problems." Science, 208: 1335-1342, 1980.

Lester, F. and J. Garofola. Mathematical Problem Solving: Issues in Research. Philadelphia: Franklin Institute Press, 1982.

Polya, G. How to Solve It. Princeton, NJ: Princeton University Press, 1957.

Ross, John A. and Florence J. Maynes. "Experimental Problem Solving: An Instructional Improvement Field Experiment." Journal of Research in Science Teaching, 20 (6): 543-556, 1983.

Descriptors--Elementary Education; *Elementary School Science; *Grade 6; Inservice Teacher Education; *Instructional Design; *Problem Solving; Science Education; *Science Instruction; *Science Programs; Teaching Methods

Expanded abstract and analysis prepared especially for I.S.E. by George I. Za'rour, American University.

## Purpose

The investigators attempted to evaluate the effect of instructional intervention to teach grade six students experimental problem solving. Based on a chronological account of what successful scientists do when designing an experiment, the domain of problem solving was broken into a set of seven skills. The first two skills in the set are: (1) to develop a focus for the investigation and (2) to establish a framework for the investigation. Both were selected in the design of the instruction, but the problem of controlling variables was addressed in the study within the context of the second skill. Performance of students in the treatment and control groups was compared.

## Rationale

A hierarchy of cognitive behaviors (levels of intellectual performance) that could reasonably be expected of grade six students was developed based on the differences between novice and expert problem solvers. For the second skill of 'establishing a framework for an investigation', the learning hierarchy consisted of four performance levels involving 11 operations as follows:

I. 1. Manipulate equipment.

II. 2. Identify a cause. 3. Select a way to measure the cause. 4. Select different amounts of the cause to test. 5. Make up a chart to record observations.

III. 6. Identify a potential effect. 7. Select a way to measure the effect. 8. Decide how many times the experiment should be done. 9. Revise the chart to record observations.

IV. 10. Make a list of other possible causes. 11. Keep other possible causes constant.

The purpose of instructional intervention is to accelerate growth of cognitive behaviors up the hierarchies with the assumption that most students will pass through the levels in the outlined sequence. In this st dy, instructional intervention involved regular teachers and intact classes rather than the practice in the studies cited in the paper in which instruction was carried out by investigators to students individually or in small groups.

## Research Design and Procedure

A pretest-posttest control group design was used for data collection. The sample consisted of 265 grade six students in 12 volunteer classrooms of one school system. Classrooms were randomly assigned to early (between first and second testing periods) or delayed (between second and third testing periods) treatment conditions of instructional intervention. Students within each class were randomly assigned to three test order groups and tested three times at three testing periods about one month apart. The first test scores were used as covariates throughout the analysis. During the interval between the first two testing periods, teachers of the delayed treatment groups continued to teach the regular science program "which included students following experimental recipes without explicit attention to design issues." The design provided an opportunity to test retention effects to the early group students one month after the instructional period.

The students were presented with baking pans and elastics of different lengths and thicknesses. After making different sounds with different elastics around pans, questions were invited and they were categorized, analyzed, and modified. The instructional intervention included identifying causes and effects in experimental questions, rewriting inadequate questions, manipulating the equipment, and teaching each of the operations like the 11 listed in the previous section. The culmination of the intervention was to introduce the idea of controlling variables. Three paper and pencil exercises were used to reinforce the performance of the operations required to complete a new experiment using the same apparatus. A fourth paper and pencil exercise required students to design an experiment to test a given hypothesis.

Teachers were prepared by the investigators for participating in the study. In a follow-up session, their perceptions of the program were elicited, and training in assessment was given. Three Test Forms (A, B and C) were prepared with six items each. Each test included two multiple-choice items and an open-ended item for each of the two skills of 'developing a focus' and 'establishing a framework'. All items whose results have been reported were of specific transfer, i.e., requiring the same skills as those taught but in a different experimental context and with different apparatus. Each multiple-choice item involved an experimental situation that required the performance of one of the two skills while the alternative responses to the item corresponded to different levels of the hierarchy of the skill--six levels for 'establishing focus' and four levels for 'establishing a framework.' The open-ended items required generation of 'sophisticated behavior.'

A students' score for each multiple-choice item corresponded to the hierarchy level chosen. Thus, the maximum score for each of the first skill multiple-choice items was 6 and the maximum score for each of the second skill items was 4. Accordingly, the maximum score for the four multiple-choice questions was 20. The maximum score for the two open-ended questions was 10, giving a total maximum score of 30 for each test form. The tests suffered from low reliability, and Test Forms A and C were not equivalent.

The instructional intervention was highly detailed with a plan to have it completed in about four to six 30-minute periods. In actuality, six to ten 30-minute periods were used reportedly "because teachers tended to have students actually conduct what were planned as paper and pencil experiments." The final exercise was not completed by a number of teachers.

Findings

The posttest scores (second test) of the early treatment groups were compared to the pretest scores (second test) of the delayed treatment groups with the first test scores used as covariates. Comparisons were made separately for each of the test order sequences (ABC, BCA, CAB) and for different parts of the tests, multiple-choice vs. open-ended. Except for one difference on the open-ended test, all the differences are in the expected direction. Comparison of posttest and pretest scores of the multiple-choice components of the tests for each treatment reflected that all student groups but one (in the delayed treatment group) performed significantly better on the posttest than they did on the pretest. On this basis, it was concluded that the instructional intervention had a beneficial effect. Comparison of posttest and retest results of the early treatment group reflected that achievement did not decline one month after the end of instructional intervention. For the open-ended instruments, almost all the findings were in the anticipated direction.

Calculating the effect size for each item of the tests and computing the average effect size for each type of item for each skill (Table VI in the report) showed that the instructional intervention was more effective in improving student performance on the skill of 'developing a focus' than on the skill of 'establishing a framework.' It also showed that the effects were more visible in the open-ended parts of the tests and larger for the early treatment group.

Interpretations

The beneficial effect of the instructional intervention is attributed to factors like practice in component operations and exposing deficiencies of cognitive strategies, thus motivating students to modify their mental operations. The relatively low size of the program effect is attributed to factors like class size, amount of treatment, and nature of practice. The investigators are of the opinion that the modest gains in experimental problem-solving that are reflected in this study can be replicated by the average teacher in the average sixth grade classroom.

## ABSTRACTOR'S ANALYSIS

The authors conclude their report of the investigation by stating that they derive from the data optimism that the students can learn to do what scientists do. This conclusion may be looked at as supporting evidence of Bruner's famous statement in The Process of Education (1960, p. 33) that "any subject can be taught effectively in some intellectually honest form to any child at any stage of development." There may be a definite value in breaking down problem solving into components and sub-components and trying to build a learning hierarchy from the parts. However, in zeroing in on two out of seven components, the investigators have not challenged the students into the complete solution of problems. There are multiple ways of solving a problem in science as in mathematics, and no support was provided for the assumption of the authors "that most students will pass through the levels in the sequence outlined..."

There are several points that are positive about the study and these are illustrated by the following:

a) The study deals with a genuine and significant problem in science education. Knowing more about the steps followed by students in solving problems could lead to better teaching. Furthermore, identifying the difficulties faced by students is likely to point to the instructional approach that may be helpful.

b) The problem is well-conceptualized and well-documented.

c) An effort was made to adapt a rigorous design--early and delayed treatment, different test forms, varying test orders, and training the teachers and test scorers.

d) A serious effort was made to approximate the natural classroom situation.

e) The investigators developed assessment tools for such investigations as reflected in another report of theirs cited in their list of references.

f) The authors recognize and list in the last paragraphs several possible sources of external invalidity.

In spite of the multitude of positive aspects, several important questions arise concerning the investigation. With respect to assessment, for example, the instruments used seem to have the following weaknesses (only the report was available to the abstractor):

a) The multiple-choice items are not of the standard type. Each option consists of cognitive behaviors classified in their totality to be at a certain cognitive level. Each option is assigned a numerical score corresponding to the level. As described in their 1983 article cited in the references, "lower levels of each hierarchy are not so much incorrect as incomplete approximation of sophisticated behavior."

b) The data were treated as if they constituted an interval scale (involving equal measurement units), which is questionable. A score of 12, for example, in the four multiple-choice questions could be a result of a variety of combinations totalling 12 such as: 5, 5, 1, 1; 3, 3, 3, 3; 5, 4, 2, 1; 1, 3, 4, 4; etc. "The test scores locate students on a scale that indicates not only how well students can perform a particular skill, but also the category of instructional acts that is likely to be most conducive to cognitive growth" (Ross and Maynes, 1983). With this description, it probably would have been more appropriate to analyze and interpret items individually rather than combined.

c) The reliabilities of the instruments are quite low. Given the type of scoring and the small number of items, this is expected. It is to be noted that the probability of getting by chance the maximum score for a framework item is 1/6 and for a focus item is 1/4.

d) With the nature and length of the options as provided in the example (p. 549), one wonders about the extent of the effect of verbal proficiency on students' performance. From another angle, the best option (level 4) in the provided example has the word "same" repeated in it three times with no occurrence of the word "sc[illegible]" which appears in the other options. It is expected that the word "same" was emphasized in the instructional intervention and, if this is true, it would have provided a hint for the reasonably test-wise student.

e) Group assessment through the paper-and-pencil test is efficient and is to be preferred to other methods such as the "think aloud" or "interview" if it realizes or approaches realizing the objectives of assessment. A major objective is to infer the nature of processes underlying the students' problem solving approach. Tests to demonstrate problem solving ability are supposed to go beyond recall or recognition of the procedures to solve a problem similar to one solved earlier.

f) The unanticipated difference between prior and pretest scores of the delayed group in the multiple-choice instruments was attributed to lack of equivalence (difference in difficulty) of Test Forms A and C. The similar unanticipated improvement of the delayed group between (Form A) and pretest (Form B) of the open-ended instrument was not explained. If it is established that two out of three Test Forms are not equivalent, then it is not justifiable to attribute differences to a genuine experimental effect. It must be admitted, however, that using class as an analysis unit tends to counterbalance the effect of non-equivalence of Test Forms; but the data for this analysis are not provided.

Some other points that can be raised about the study are as follows:

a) The use of the same variable as a covariate and as a dependent variable is highly questionable (Evans and Anastasio, 1968).

b) The sampling unit is the class whereas the analysis unit is the student. This has a large effect on "significant" results because of the difference in df. The authors report reexamination of the data using class rather than student with similar results. It is not clear whether this similarity includes "significance."

c) Little information is available in the report about the teaching-learning situation engaged in by the students of the control group. In other words, it is not very clear what the control group did during the experimental period. It is very likely that differences between groups emanated from the "opportunity-to-learn" variable rather than by transfer. A strong bias against the control group is possible because of the kinds of experience they did not have.

d) The teacher variable does not seem to have been adequately controlled. In spite of the detailed instructions, the teachers did not follow the plan closely and required 6 to 10 periods instead of 4 to 6, and there was a variation in what was accomplished in the different classes.

e) The article is difficult to decipher.

With respect to further research, analysis of individual items and choices may reveal certain patterns worthy of further investigation or relevant to future teaching of similar topics. The order of complexity within hierarchies may be established through Guttman analysis.

To conclude, Ausubel and Robinson (1969, p. 505) state that "...a true problem would not involve a well-defined or invariable sequence of transformations which had been practiced to the point where they could be run off routinely as a straightforward application or mere recollection of a rehearsed solution." The reported insignificant effect on general transfer as compared to specific transfer tends to reflect a degree of mastery of facts and techniques rather than structure. The authors started with the observation that the often-sought objective of ability to solve problems is not adequately realized. Their instructional intervention was an attempt to improve the situation. On the basis of this study, as reported, and considering its methodological pros and cons, one is tempted not to give much weight to the "optimism that the slogan of the curriculum reform movement in science education can be realized and that students can learn to do what scientists do." (p. 555)

## REFERENCES


Ausubel, David P. and Floyd G. Robinson. School Learning--An Introduction to Educational Psychology. New York: Holt, Rinehart and Winston, Inc., 1969.

Bruner, Jerome. The Process of Education. Cambridge: Harvard University Press, 1960.

Evans, S. H. and E. J. Anastasio. "Misuse of Analysis of Covariance when Treatment Effect and Covariance are Confounded." Psychological Bulletin, 69: 225-234, 1968.

Ross, J. A. and F. J. Maynes. "Development of a Test of Experimental Problem-Solving Skills." Journal of Research in Science Teaching, 20 (1): 63-75, 1983.

RESPONSES

57

IN RESPONSE TO THE ANALYSIS OF

Ross, John A. and Florence J. Maynes. "Experimental Problem Solving: An Instructional Improvement Field Experiment" by George I. Za'rour. Investigations in Science Education 12 (3): 42-50, 1986.

John A. Ross
The Ontario Institute for Studies in Education

Za'rour briefly identifies some strengths of the study reported by Ross and Maynes (1983). Most of his energy goes toward itemizing its methodological shortcomings. We agree in large part with the statement of methodological deficiencies, although we have some reservations about some of the specifics raised in the review. Where we part company with the reviewer is at the conclusion. Za'rour claims that not much weight should be attached to the findings: we think otherwise.

Za'rour is on strong grounds when he cites inadequacies in the multiple choice instruments used in the study. We note, however, that similar results were generated by the open-ended items. We agree with the reviewer's concerns and for much the same reasons we now rely exclusively upon open-ended instruments in our investigations (see, for example, Ross and Robinson, in press). Incidentally, Za'rour suggests that it would have been more appropriate to analyze items individually rather than aggregating the items. Actually we did both. Only the combined scores were reported due to space limitations. For the same reason the class means analysis was not included.

Other concerns raised by Za'rour are not as straightforward. He implies that the use of pretest scores as a covariate is inappropriate but provides no rationale, other than an incomplete reference, for his position. Similarly he cites as a weakness the analysis of student scores when the class is the sampling unit. In each case the reviewer has simply expressed a methodological preference: there is no consensus among researchers on these issues. With respect to the latter, for example, Hopkins (1982) provides a literature review and empirical demonstration to show that the "recommendation to use group means when there may be nonindependence among observational units is unnecessary, unduly restrictive, impoverishes the analysis and limits the questions that can be addressed in a study" (p. 5).

The reviewer observes that insufficient information is given on the instructional program provided to control group students and that the researchers did not adequately control for the teacher variable. Both criticisms are valid and the original article suggested other ways in which the study failed to control for factors that might affect the results. This is hardly surprising: it is impossible to maintain a tightly controlled design in a field setting (Cronbach, 1982).

The obvious solution to these methodological problems is to avoid research in field settings. But this solution creates another problem of equal seriousness. The well controlled study contaminates natural processes; the lab study lacks ecological validity. The end purpose of research in science education is not to promote learning when students are under the control of researchers, although this can be a worthy intermediary goal, but to promote learning when students are under the control of ordinary teachers in ordinary classrooms.

Clearly what is required is both types of studies. Yet in the past the lab study has been the overwhelming preference of researchers. For example, Sneider et al (1984), in a review of published studies attempting to teach controlling variables, reported that only one study (Ross and Maynes, 1983) had, at the time of the review, used the students' usual teachers as instructors. All the others used researchers in this role.

This is the value of the study. It provides a needed balance to the literature. A more apporpriate conclusion to Za'rour's review would be a statement which recognizes that the methodological shortcomings create uncertainty about the findings, that one could not therefore make recommendations to teachers based on this investigation alone, but together with previous investigations in lab settings, it provides "grounds for optimism that the slogan of the curriculum reform movement in science education can be realized and that students can learn to do what scientists do."

REFERENCES


Cronbach, L. S. Designing Evaluations of Educational and Social Programs. San Francisco, CA: Jossey-Bass, 1982.

Hopkins, K. D. "The Unit of Analysis: Group Means Versus Individual Observations." American Educational Research Journal, 19 (1): 5-18, 1982.

Ross, J. A. and F. J. Maynes. "Experimental Problem-Solving: An Instructional Improvement Field Experiment." Journal of Research in Science Teaching, 20 (1): 543-556, 1983.

Ross, J. A. and F. G. Robinson. "The Use of Rule Structures in Teaching Experimental Design to Secondary School Students." Science Education, in press.

Sneider, C., K. Kurlick, S. Pulos and A. Friedman. "Learning to Control Variables with Model Rockets: A Neo-Piagetian Study of Learning in Field Settings." Science Education, 68 (4): 456-486, 1984.

IN RESPONSE TO THE ANALYSIS OF

Tamir, Pinchas. "Teachers' Self Report as an Alternative Strategy for the Study of Classroom Transactions" by Constance M. Perry. Investigations in Science Education 12 (3): 11-18, 1986.

Pinchas Tamir
School of Education and Israel Science Teaching Center
Hebrew University, Jerusalem

I am glad to have the opportunity to clarify a number of issues raised in the review of my paper.

Figure 1 presents the SLRF (biology) as requested by the abstractor. It may be noted that somewhat different forms were used with physical science classes. For example, different items were listed in item 27 regarding the kind of lab equipment used. However, most items apply to all science areas. Many of the questions raised may be answered by examining Figure 1.

1. It is true that any instrument is biased by the kind of information it collects. However, the open-endedness of item 28 (description of lesson), as well as the opportunity to provide "other" responses in many items, reduces the bias considerably, although of course the bias of the observer remains.
2. A teacher can complete the SLRF in about 20 minutes. Teachers often agree to invest the time especially if they are convinced that the study is important.
3. Science educators who established validity based their judgment on comparing the items of SLRF as well as the categories of the LESSON ANALYSIS SCHEME (See Figure 2) to their experience and familiarity with science classes (face validity).
4. As to reliability, the observers indeed completed the forms d ring the lesson as well as after the lesson, depending on the items. or example, the score item No. 28 - lesson description - was completed during the lesson, while teachers' comments (e.g. items 35, 36) or general information (items 37-43) were completed after class.

5. The 56 percent agreement reached with regard to students' satisfaction and enjoyment is indeed the lowest level.
6. In the junior high, 38 percent of the lessons were planned with the aid of the teacher's guide (item 16) and 65 percent followed the sequence recommended in the student's text (item 18).
7. Inquiry was assured in two ways: a) the teachers' and observers' general assessment (item 30); and b) by responses to other items such as 28 (See Figure 2), or 26 and 27.

I am pleased with the general conclusions of the abstractor and look forward to other researchers who would use SLRF.

FIGURE 1

STRUCTURED LESSON REPORT FORM (BIOLOGY)

1. RECITATION  2. LAB  3. INTEGRATED LAB/RECITATION

11. Lesson topic
12. Number of students in class
13. Title of textbook used
14. Title of lab manual used
15. If lab exercise not taken from lab manual list its source:__________

________________________________________

16. Did you use the teacher guide?
17. Did you modify the recommended lesson presentation? If yes - how?

________________________________________

18. Did you follow the sequence of lessons as suggested by the program? If not, give your reasons:____________________

________________________________________

19. What kind of organisms were dealt with? a) plants b) animals c) humans d) micro-organisms e) the cell f) no special organism
20. What areas were dealt with? a) physiology b) biochemistry c) ecology d) genetics e) evolution f) taxonomy g) morphology h) behavior i) other - which?____________________

________________________________________

21. GRADE LEVEL OF CLASS:

22. Students characterization: a) slow learner b) mixed ability c) average d) above average e) other:__________________________

23. THE CLASSROOM: a) regular b) lab c) other__________________________

24. Sitting arrangement: a) individual desks b) two per desk c)other

25. Work pattern: a) whole class b) pairs c) 3 - 4 in each group d) individual

26. Instructional aides used: a) chalkboard b) overhead projector c) slide projector d) super 8 loop projector e) 16mm film projector f) pictures, charts and models g) plants h) live animals i) preserved animals j) other__________________________

27. Lab equipment used: a) microscope b) binocular c) scale d) dissecting set e) centrifuge f) spectrophotometer g) other__________________________

28. Description of lesson:__________________________

29. Homework assigned: a) NONE b) questions from textbook c) teacher's questions d) reading from textbook e) reading from other sources h) written summary/lab report i) performing an independent investigation j) other:__________________________

30. TO WHAT EXTENT WAS THE CLASS INQUIRY ORIENTED? a) not at all b) slightly c) moderately d) to a large extent

31. DID THE STUDENTS ENJOY THE CLASS? a) No b) I don't know c) Yes

32. What is the basis for your answer to item 31? a) a lot of discipline problems b) I talked with students after class c) I noticed their responses in class d) other:__________________________

63

33. Did you enjoy the class? a) No b) not sure c) Yes

34. What made you enjoy/not enjoy the class? a) discipline problems b) my objectives were not achieved at all c) my objectives were partially achieved d) my objectives were fully achieved e) student participation was good f) other: ______________________

______________________________________________

35. How similar is this lesson to other lessons in this class?
a) not similar b) somewhat similar c) very similar

36. How similar is this lesson to lessons you teach in other classes?
a) not similar b) somewhat similar c) very similar

37. DATE OF CLASS: ______________________

38. NAME OF SCHOOL: ______________________

39. NAME OF TEACHER: ______________________

40. Number of years in teaching: ______________________

41. College degree: ______________________

42. Major field of study: ______________________

43. Kind of preservice education: ______________________

LESSON ANALYSIS SCHEME

| ACTIVITY | WHEN PERFORMING | | | WHO PERFORMED | | |
|---|---|---|---|---|---|---|
| | BEGINNING | MIDDLE | END | TEACHER | TEACHER AND STUDENTS | STUDENTS ONLY |
| Stating a Problem | | | | | | |
| Demonstrating | | | | | | |
| Reviewing Literature | | | | | | |
| Explaining | | | | | | |
| Reading Homework | | | | | | |
| Writing in Notebooks | | | | | | |
| Reading Textbook | | | | | | |
| Reading Other Materials | | | | | | |
| Discussing | | | | | | |
| LAB ONLY | | | | | | |
| Pre Lab Discussion | | | | | | |
| Planning Experiment | | | | | | |
| Performing Experiment | | | | | | |
| Pooling Results | | | | | | |
| Post Lab Discussion | | | | | | |