DOCUMENT RESUME

ED 222 344 SE 039 393

AUTHOR Blosser, Patricia E., Ed.; Mayer, Victor J., Ed.
TITLE Investigations in Science Education. Vol. 8, No. 4.
INSTITUTION ERIC Clearinghouse for Science, Mathematics, and Environmental Education, Columbus, Ohio.; Ohio State Univ., Columbus. Center for Science and Mathematics Education.
PUB DATE 82
NOTE 83p.
AVAILABLE FROM Information Reference Center (ERIC/IRC), The Ohio State Univ., 1200 Chambers Rd., 3rd Floor, Columbus, OH 43212 (subscription $6.00, $1.75 single copy).
JOURNAL CIT Investigations in Science Education; v8 n4 1982

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS *Academic Achievement; Cognitive Development; *College Science; Elementary School Science; Elementary Secondary Education; Higher Education; Individualized Instruction; Inquiry; Measures (Individuals); *Preservice Teacher Education; Process Education; Science Education; *Science Instruction; *Secondary School Science; Student Characteristics; Teacher Characteristics; *Teaching Methods
IDENTIFIERS *Science Education Research

ABSTRACT

Abstractor's analyses of 12 science education research studies focusing on science instruction are presented. These include: (1) use of Learning Environment Inventory in junior high school classrooms organized for individualized instruction; (2) microteaching and strategy analysis as used in a science methods course; (3) comparison of two different teaching methods in a college physics course for non-science majors; (4) comparison of effects of didactic and inquiry teaching in cooperative and competitive settings; (5) interaction of teaching method, level of student intelligence and sex in problem-solving tasks; (6) comparison of lecture and audio-tutorial instruction on acquisition of science process skills by preservice teachers; (7) modeling as a method for acquiring teaching skills; (8) use of extra credit opportunities by college freshmen enrolled in a large-enrollment biology course; (9) use of filmed experiments as an alternative to study-centered laboratory work in chemistry; (10) development of an instrument for assessing scientific literacy of secondary school students; (11) an instrument for use in distinguishing between relatively concrete and relatively formal levels of logical development; and (12) students' reasoning abilities, achievement in high school chemistry, and misconceptions they held concerning chemical equilibria. Author's responses to abstractor's analyses of the latter three studies are also presented. (Author/JN)

ED222344

INVESTIGATIONS IN SCIENCE EDUCATION

National Association for Research in Science Teaching

ERIC® Clearinghouse for Science, Mathematics,
and Environmental Education

INVESTIGATIONS IN
SCIENCE EDUCATION

Volume 8, Number 4, 1982

Published Quarterly by

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

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

SE039393

Volume 8, Number 4, 1982

NOTES FROM THE EDITOR: -

The theme of Volume Number 8, Issue 4 of INVESTIGATIONS IN SCIENCE EDUCATION is one emphasized in an earlier issue--Instruction. Fraser reported on the use of the Learning Environment Inventory in junior high school classrooms organized for individualized instruction. Yeany studied the effects of microteaching and strategy analysis as used in a science methods course. Renner and Paske compared two different teaching methods as used in a college physics course for nonscience students. Tjosvold and his colleagues compared the effects of didactic and inquiry teaching in cooperative and competitive settings. Ryman examined the interaction of teaching method, level of (student) intelligence and gender in problem-solving tasks. Romaro compared lecture and auto-tutorial instruction on the acquisition of science process skills by preservice teachers. Santiesteban and Koran examined modeling as a method for acquiring teaching skills. Canary et al. investigated the use of extra credit opportunities by college freshman enrolled in a large enrollment biology course. Ben Zvi et al. studied the use of filmed experiments as an alternative to study-centered laboratory work in chemistry.

In the "Critiques and Responses" section, Rubba and Anderson reported on the development of an instrument for assessing the scientific literacy of secondary school students. Wollman described an instrument for use in distinguishing between relatively concrete and relatively formal levels of logical development. Wheeler and Kass looked at students' reasoning abilities, their achievement in high school chemistry, and misconceptions they held concerning chemical equilibria.

Patricia E. Blosser
Editor

Victor J. Mayer
Associate Editor

INSTRUCTION

Fraser, Barry J. "Measuring Learning Environment in Individualized Junior High School Classrooms." Science Education, 62(1): 125-133, 1978.

Descriptors--*Classroom Environment; *Classroom Observation Techniques; *Educational Research; *Individualized Instruction; Junior High Schools; *Measurement Instruments; Science Education; *Secondary School Science

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Linda R. DeTure, Rollins College.

Purpose

The two purposes of this study were to adapt Walberg's Learning Environment Inventory (LEI) for use in individualized junior high school science classes and to validate the instrument by administering the modified version to a sample group of seventh graders.

Rationale

The underlying premise of this study was aimed toward finding a more valid means of measuring the total learning environment. Increasingly the goals of innovative science curricula are to enhance the complete learning environment which encompasses the instructional processes as well as pupil affective and cognitive outcomes. However, measuring the effectiveness of the total environment has been somewhat of an obstacle. A common approach, direct observation or low inference measures, has been criticized for high cost and because the results frequently account for only a small amount of variance in student learning.

As an alternative, Fraser cites a number of studies in which high inference measures, based on the pupil's perception of the learning environment, have been shown to be reliable, economical and good predictors of student learning outcomes. Some limitations of past research with high inference measures are that the instruments have been designed only for certain contexts. For example, the LEI was written for the senior high school level, making its reading level too high to be suitable for use in the junior high school. Since research with the original LEI has demonstrated both its reliability and predictability, Fraser aimed to modify Walberg's LEI for individualized junior high science classes and to alter the readability to the appropriate level.

The theory and research involved in the validation of the LEI are examined in some detail and an extensive reference section is included.

Research Design and Procedures

At the suggestion of Anderson and Walberg, the researcher excluded several of the fifteen LEI scales that had little relevance to individualized classrooms. A shortened instrument that could be administered in one classroom was the result. The four-point Likert response format remained the same. A panel of educators decided which scales should be omitted. The following eight scales were retained: Diversity; Speed; Environment; Goal Direction; Satisfaction; Disorganization; Difficulty; Completeness. Additionally, a new scale was added to tap the dimension of individualization. Many of the items were reworded in a way to lower the reading level but to keep the wording as similar to the original items as possible.

The modified version to the LEI was administered to 541 science students in 20 classes, ten using conventional science curricula materials and ten using the individualized materials of the Australian Science Education Project, ASEP.

Findings

The modified instrument was validated in terms of three statistical criteria: interna consistency, discriminant validity (to measure correlations between tne scales in the battery), and sensitivity. The validation techniques were previously reported in another paper. Item indicies of the three criteria were used to identify faulty items. Upon removal of certain items, 55 remained in the nine-scale battery and the data were reanalyzed. An α reliability coefficient was used as an index of internal consistency; scale intercorrelations were the indices of the discriminant validity; and student total scores were used for sensitivity.

The reliability ranged from 0.50 to 0.80 with a median of 0.63. The author considered these sufficiently high to indicate internal consistency for each scale. The internal correlations between scales ranged from 0.00 to 0.48 with a median of 0.23. The low correlations indicated to the researcher that distinct constructs exist within the scales. The range of student scores covered most of the score range for all scales and thus the instrument's sensitivity was considered satisfactory.

The two groups of classes were chosen to be as comparable as possible except for the curriculum materials. The two groups were compared on each of the nine scales using multiple regression techniques. The results of these analyses were also reported elsewhere and were not given in this paper. Fraser reports that the analyses revealed that the ASEP students viewed their classes significantly different from the conventional curriculum material students on three of the scales: Environment, Satisfaction, and Individualization.

Interpretations

Because students in the individualized classes rated their classes higher than did the control group in terms of environment (availability of materials and resources), satisfaction and individualization, Fraser feels that the evidence provides some support for the usefulness of the instrument for measuring the learning environment in individualized science classrooms at the junior high school level.

ABSTRACTOR'S ANALYSIS

This paper addresses an area of science education that deserves considerable attention, namely how to measure the effects of all those intangible goals of a total curriculum program. Conventional means of testing learning outcomes have not proved to be good for evaluating certain aspects of the curriculum program. For example, the effectiveness of laboratory as a learning tool in the science classroom is always a good source for debate whether considering cognitive or affective gains. Perhaps what is happening with the laboratory question and with the other instructional processes is that the total learning environment is not being considered. The measurements being used probably do not provide sufficient data to make critical decisions about the effectiveness of a particular curriculum. Studies like this one seek to further examine and refine high inference measures provide an important direction for those searching for answers involving the total learning environment.

The author has a straightforward writing style that made the paper flow and was easy to decipher. Also, he provided a wealth of background references for anyone interested in pursuing this line of study. However, it appears that he has combined the research of two previously published papers into this paper which is more a summary of research than a research report. In the comparison of the two groups, the findings are summarized and interpreted, but no statistical data are given. Thus, there is no opportunity for the reader to make an independent interpretation because the original journal in which the statistics were presented is not widely available.

The data tables that are reported in this article are related to the validation of the instrument. In that regard a couple of questions are raised. Concerning readability of items, no mention was made of a reading test being used to determine the reading level of the original instrument. In a study in which modifying the reading level is one of the primary goals, it would be appropriate to know the reading level of the LEI by subjecting it to a reliable reading test. The panel of experts could then make the adjustments in the instrument until the correct reading level was reached as measured again by a reading test.

Some of the scales in the battery have so few items (Diversity-4) that the internal consistency may be weaker or stronger than is indicated by the reliability coefficients. Another question is raised relating to the sensitivity of the instrument. Although most scales show distribution along all point ranges, several appear to be skewed in one direction. Although it is difficult to determine from the tables, it appears that the instrument may need some revision and further validation. One suggestion is that a faulty item could be reworded rather than removed, thereby increasing the number of items in the abbreviated scales.

In conclusion, it is likely that the two previously reported papers contain the information that a researcher would want to review before attempting to use the Modified Learning Environment Scale. This paper simply suggests that the high inference instrument has some utility in addressing the questions concerning the total learning environment in junior high school individualized science classrooms.

REFERENCES

Bates, Gary C. "The Role of the Laboratory in Secondary School Science Programs," in What Research Says to the Science Teacher, Volume I edited by Mary Budd Rowe. National Science Teachers Association, 1978.

Fraser, B. J. "Selection and Validation of Attitude Scales for Curriculum Evaluation." Science Education, 61(3):317-330, 1977.

Fraser, B. J. "Pupil Perceptions of the Climate of ASEP Classrooms." Australian Science Teachers Journal, 22: 127-129, 1976.

Yeany, Russell H. "Effects of Microteaching with Videotaping and Strategy Analysis of Preservice Science Teachers." Science Education, 62(2): 203-207, 1978.

Descriptors--College Science; *Educational Research; Higher Education; *Microteaching; *Preservice Education; *Science Education; Science Teachers; Teacher Education; *Teaching Skills

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Hans O. Andersen, Indiana University.

Purpose

The investigator's purpose in conducting this study was to assess the effects of microteaching with videotape playback and strategy analysis on the teaching strategies selected by preservice secondary science teachers.

Rationale

Microteaching with videotape playback and strategy analysis has been used extensively by methods instructors. It is assumed that the simulated experience so provided will help students develop appropriate teaching strategies when the teaching is coupled with student self-analysis of performance and instructor feedback. In this study the investigator used the Teaching Strategies Observation Differential (TSOD) for student self-analysis and also to measure teaching behavior changes occurring during the period of the study. The teaching behaviors emphasized in this study were classified as the inductive-indirect teaching behavior. The investigator chose to study these behaviors after reviewing the research on science achievement and attitude toward science. These behaviors are frequently found to cause, or at least be correlated with, science achievement and positive attitude development. The investigator attempted to determine if his instruction did influence students to use these good "research-supported" behaviors.

Research and Design Procedures

The subjects of the study were undergraduate science majors at Southern Illinois University, Carbondale, Illinois. They were enrolled in the preservice education program. They had, for the most part, completed their science requirements and would be student teaching during the next term.

The three treatments were randomly assigned to three intact sections scheduled at different times to eliminate student interaction. All subjects taught a pretreatment and posttreatment peer group lesson. Both lessons were videotaped. The videotaped first lesson was used as part of the instruction and to provide data for the later occurring covariate analysis. The treatment levels were varied in the following manner:

Level I. Private review of prelesson without any guidelines or instruction.

Level II. Private review of prelesson after instruction in using the Teaching Strategies Observation Differential (TSOD). The students' coding sheets were collected but not discussed or evaluated.

Level III. This treatment consisted of the level two strategy plus an additional session with the instructor in which both viewed the tapes and agreed upon the classification of the 30-second intervals of behavior: "The main task was to systematically define type of strategy exhibited in the lesson."

A trained rater was employed to use the TSOD to analyze the pre- and post treatment tapes. A Pearson correlation indicated a rate-rerate reliability estimate of .93.

The TSOD was used to measure the teaching style on the continuum from expository-direct to inductive-indirect. Pretreatment scores were used as the covariate in this analysis. ANCOVA procedures were used in order to remove possible selection bias resulting from the use of intact groups and to increase the statistical power of the hypothesis testing. The investigator reported a prior decision to test all hypotheses at the 0.10 alpha level and identified treatment level I as the control group for the post hoc analysis. The Dunnett test and Newman-Keuls technique were employed for the post hoc analysis.

Findings

1. Significant differences among the means of the three treatment groups existed. ANCOVA, ($p < 0.001$).
2. Level II and Level III treatment subjects used significantly more indirect teaching behaviors than Level I subjects. Dunnett test, ($p < 0.01$).
3. Level II and Level III treatment subjects used significantly more indirect teaching behaviors than Level I subjects. Level III subjects used significantly more indirect teaching behaviors than did Level II subjects. Newman-Keuls, ($p < 0.05$).

Interpretation

Students who used the TSOD to analyze their teaching used more indirect teaching behaviors than did their untrained counterparts. Furthermore, when a second viewing of the lesson with instructor present was employed, the students' use of indirect influence in their post lessons was increased even more. The idea that instructor provided feedback will positively influence teaching performance of preservice teachers was supported.

ABSTRACTOR'S ANALYSIS

The investigator's interest in basing methods instruction on research in science education is noteworthy. His efforts to evaluate the effectiveness of his instruction in terms of student achievement instead of student testimonial similarly deserves recognition. One can hardly expect one's students to worry about their students' achievement without exhibiting similar concerns.

In this study the investigator used three treatment levels. The first treatment level, the control group, received minimal instruction. The students were only told to look at their videotapes. The second treatment subjects were provided instruction in using systematic observation. The fact that these students used significantly more indirect teaching behavior illustrated the value of systematically studying teaching behavior. The untrained Level I students did not see the hierarchically arranged TSOD or calculate where they fit on the direct-indirect continuum. One would expect that they would not use as great a variety of influence and one would expect to discover that trained students used more indirect behavior. Treatment Level III was an extension of Level II. The instructor essentially provided more training on the TSOD. Again the post treatment subjects used significantly more indirect behavior than their less instructed counterparts. The conclusion that inserting the instructor caused the improvement is certainly one that I would endorse! I too want to feel that my presence makes significant difference. However, the instructor claimed that his only role was clarification. That is, reinforcement was not provided to students for using indirect behavior. Hence, one might pose as alternate hypothesis that inserting more paper and pencil instruction on the TSOD would have an effect equal to that provided by the instructor.

Post microteaching teacher-student conference techniques have most frequently been nondirective as was the case in this study. Forcing the student to make the decision has many merits; however, it is possible that students should be given more direction in the earlier phases of their training.

Stolurow's classic article "Model the Master Teacher or Master the Teaching Model" (1965) advances the argument for mastering the model. Yeany, in the preface to this article (1978), states, "If we intend to base our science teacher training activities on the results of available research (which at times may be scant, but promising), then it would be appropriate to encourage and train teachers in the use of a range of teaching strategies which include inductive-indirect approaches." Yeany thus endorsed the Stolurow argument and put it into practice by training students to use the hierarchically arranged TSOD. In spite of the conference technique used, most students undoubtedly got the message that the indirect approach was the desired approach.

Research in available teaching models also has considerable support (Okey et al., 1973; Joyce and Weil, 1972; Eggen et al., 1979). Then there is the issue of accountability. Should we continue the nondirective approach, assuming that better teachers will evolve because they are increasingly more self-critical? Or, should we say to our students, "Here are four or five or six teaching models. When you have demonstrated mastery of them, we will recommend you for certification." The nondirective approach has not seemed to produce a cadre of excellent constantly improving teachers. Will a directive approach be better? Is a hybrid needed?

REFERENCES

Eggen, P. D.; D. P. Kauchak; and R. J. Harden. Strategies for Teachers. Englewood Cliffs, NJ: Prentice Hall, 1979.

Joyce, B. and Weil, J. Models of Teaching. Englewood Cliffs, NJ: Prentice Hall, 1972.

Okey, J. R.; H. O. Anderson; and J. L. Brown. "Competencies for Science Teachers." in Turner, R. L. A General Catalogue of Teaching Skills. Syracuse University: Multi State Consortium on Performance-Based Teacher Education, December 1973.

Stolurow, Lawrence. "Model the Master Teacher or Master the Teaching Model," in Krumboltz, J. D. Learning and the Educational Process. Chicago: Rand McNally, 1965.

Yeany, R. H. "Effects of Microteaching with Videotaping and Strategy Analysis on Teaching Strategies of Preservice Science Teachers." Science Education, 62(2): 203-207, 1978.

Renner, John W. and William C. Paske. "Comparing Two Forms of Instruction in College Physics." American Journal of Physics, 45(9): 851-859, September 1977.
Descriptors--*Achievement; *College Science; *Educational Research; Higher Education; *Instruction; Physical Sciences; *Physics; Science Education; Teaching Methods

Expanded abstract and analysis prepared especially for I.S.E. by Frank A. Smith, Jr., West Chester State College.

Purpose

This investigation compares two types of teaching methods identified as "concrete instruction" and "formal instruction." The authors hypothesized that "the amount of learning that takes place in a classroom is a function of the teaching method employed." Although not explicitly stated, one can infer that the following more specific hypotheses were tested:

I. Students taught by the concrete method will score higher on a physics content examination than students taught by the formal method.

II. Students taught by the concrete method will show larger gains on five selected tasks to measure problem-solving ability than will students taught by the formal method.

III. Students taught by the concrete method will show larger gains on three Piaget tasks than will students taught by the formal method.

The study also assessed the students' satisfaction with the method of instruction.

Rationale

Recent Piagetian research indicates that many college students are at a stage of intellectual development identified by Piaget as concrete operational while some have reached the higher formal operational stage. In many college physics courses the concepts presented and the manner in which they are presented assumes that the students have reached the formal operational stage. The fact that many students, particularly nonscience students, have not reached this formal stage has implications for college physics teaching. The authors postulate that the amount of learning that takes place in a physics course for nonscience students can be increased by using a concrete method of instruction rather than a formal method of instruction.

Research Design and Procedure

The study compares the performance of students in three concrete instruction groups, the experimental groups, to the performance of students in one formal instruction group, the control group. The students were nonscience students taking a one-semester course in introductory physics at the University of Oklahoma. The dependent variables under investigation were content understanding, as measured by performance on a 20-item examination, and performance on six different instruments designed to measure outcomes other than content understanding. The six instruments were:

1. The Fuller Task: A task designed to measure the ability to utilize ratio and proportion.
2. The Five Ratio Tasks: A six-item test to measure the students' ability to utilize ratios.
3. The Karplus Ratio Task: A task to measure proportional thinking ability.
4. The Karplus Islands Puzzle: A task to measure logical thinking ability.
5. The Watson-Glaser Critical Thinking Test: A 100-item test to measure critical thinking ability.
6. Three Piaget-designed Tasks: The tasks were the conservation of volume, equilibrium in the balance, and the separation of variables. These three tasks were used to classify students as to concrete and formal thought levels.

The basic research design used was a nonequivalent control group design where both experimental and control groups were given pretests and posttests on each of the measurements made except for the measure of content understanding where no pretest was given.

The students were allowed to enroll in any one of five sections of two courses and were allowed to transfer from section to section during the first two weeks of class. The sample sizes differ from measurement to measurement. For the evaluation of content understanding there were 11 students in the formal group and a total of 52 students in the three concrete groups. The sample sizes vary for the other measurements but were approximately the same.

For the measurement of content understanding the mean score on the content test for the formal group was compared to the mean score on the content test for each of the concrete groups by means of the t-test. Mean scores of the concrete groups were also compared to each other by means of the t-test. For the measurements using the Watson-Glaser instrument, the Fuller Task, the Five Ratio Tasks, the Karplus Ratio Task, and the Karplus Islands Puzzle the percentages of students showing gains, or losses, from pretest to postest were compared by means of a bar graph. For the Piaget-designed Tasks, point values were assigned to the Piaget stages IIA, IIB, IIIA, and IIIB and the total changes in the Piaget levels for the formal group students were compared to the total change of ten randomly assembled groups drawn from the concrete instruction groups. Also, the percentages of students making a change in Piaget levels were compared.

Findings

On the measure of content achievement the authors found that each of the concrete instruction groups had a higher mean score on the content examination than the formal group. The levels of significance generated from the t-test were .08, .06, and .0005. On the Karplus Islands Puzzle, the Watson-Glaser instrument, and the Five Ratio Tasks, the concrete groups made greater gains and experienced smaller losses than did the formal group. On the Fuller Task there was little difference between the concrete and formal groups. The data from the Karplus Ratio Task was dropped from consideration because the sample size of the formal group (7) was too small on this measure. The Piaget-designed tasks results indicated that the total changes in Piaget levels were the same for both groups. However, for the students in the formal instruction group those students who changed Piaget levels (38 percent) the majority (78 percent) did so by changing from formal A level to formal B level. In the concrete groups the tendency was for students at concrete levels to move to formal levels or in the direction of formal levels.

The results of the questionnaire indicated that students were satisfied with the concrete instruction and dissatisfied with the formal instruction.

Interpretations

From these results the authors concluded the following:

1. Students experiencing concrete instruction achieve higher scores on examinations dealing with physics content than do students experiencing formal instruction.
2. Concrete instruction promotes students' problem-solving abilities better than does formal instruction.
3. Concrete instruction promotes intellectual development at both the concrete and formal levels while formal instruction advances the intellectual development of only those students who have entered the formal operational stage.
4. Students are happier with concrete instruction than they are with formal instruction.

ABSTRACTOR'S ANALYSIS

The results of this investigation would seem to have important implications for the teaching of introductory physics courses to nonscience students. The study indicates that concrete methods of instruction produce greater content knowledge and increased problem-solving ability than do formal methods of instruction. There are, however, a number of questions about the study which bear upon its validity and generalizability. The questions which seemed important are summarized below.

Sample Size and Selection

A. The report mentions that the sample was composed of five sections of two courses but only four sections, three concrete instruction groups and one formal instruction group, are dealt with in the report. It is not clear what happened to the fifth section.

B. The formal instruction group, which serves as the control group in the experiment, was very small. The number in the formal group ranges from 11 on the content achievement measure down to seven for the Karplus Ratio Task. The total number of students in the three control groups was about 52. There is no mention in the report why it was not possible to have two formal groups and two concrete groups and a more even distribution of students.

C. Students enrolled in the two courses were allowed to transfer from section to section during the first two weeks of the semester. The authors report that the Watson-Glaser instrument was administered on the first day of class and the Piaget Tasks were administered during the first two weeks of class, but they do not mention how the data were handled for those students who may have switched from a formal to concrete group, or vice versa, during the first two weeks. Also, could this liberal transfer policy result in students selecting the mode of instruction that they felt most comfortable with?

The Instruments

The content evaluation instrument was a 20-item free-response examination. The questions were based on a list of concepts that were taught to both the formal group and the concrete groups. The question arises as to what fraction of the total number of concepts in each group does this common list represent. It is possible for the total number of concepts taught in each group to be greatly different and, if this were so, a test on the concepts taught in common might be biased in favor of the group studying the fewer number of concepts in more depth. Also, the actual writing of this test was done by the instructors teaching the concrete groups. It is also not clear from the report who administered the tests and the tasks. Were both formal instructor and concrete instructors the administrators, or was it someone else, or some combination?

Possible Pretest-Posttest Effects

The authors felt that the process of repeating the tests was not a factor in the improvement of the students' performance. However, Lawson, Nordland, and DeVito (1974) report significant posttest gains on some similar Piaget tasks with no intervening treatment between pretest and posttest.

The Instructional Methods

The concrete instructional method is described in some detail with respect to textbook used, classroom procedures, and examples from the textbook. The description of the formal instructional method is less detailed. There is no mention of the textbook used and the classroom procedure is described as a traditional lecture-demonstration method. The authors have chosen to call these methods "concrete" and "formal" but from the descriptions they could just as well have been called "inquiry or laboratory based" and "traditional." In fact, the study is more closely aligned with inquiry vs. traditional teaching method research than with Piagetian research.

In summary, a replication of this experiment under more closely controlled conditions would be desirable. The authors' findings are important enough and exciting enough to demand verification by other investigators. If such a replication is undertaken and if the investigators have the luxury of freedom in experimental design, they mignt consider a four-group design with one experimental group pretested and the other not and one control group pretested and the other not. Perhaps two instructors could be employed with each instructor teaching one control group and one experimental group, one of which is pretested and the other not.

REFERENCE

Lawson, Anton E.; Floyd H. Norland; and Alfred DeVito. "Piagetian Formal Operational Tasks: A Crossover Study of Learning Effect and Reliability." Science Education, 58: 267-76, 1974.

Tjosvold, Dean; Paul M. Marino; and David W. Johnson. "The Effects of Cooperation and Competition on Student Reactions to Inquiry and Didactic Science Teaching." Journal of Research in Science Teaching, 14(4): 281-288, 1977.

Descriptors--Achievement; Attitudes; *Didacticism; Educational Research; Elementary Education; *Elementary School Science; *Inquiry Training; *Instruction; Questioning Techniques; Science Education; *Teaching Methods

Expanded abstract and analysis prepared especially for I.S.E. by Lowell J. Bethel, The University of Texas at Austin.

Purpose

The purpose of the study was to investigate the effect of traditional didactic teaching vs. inquiry teaching on students who either compete or cooperate with each other. Specifically the study compared the students on these four factors:

1) The acceptance of the teaching method

2) Students' approval of the teacher

3) The experience of peer support

4) Students' belief that they have learned

There were four hypotheses tested.

Rationale

Inquiry teaching has not been widely accepted by teachers even though a strong theoretical basis for it exists. In addition there also appears to be empirical support for its effectiveness as a teaching strategy. Many modern school curricula (e.g. BSCS, SCIS, ESS, MACOS, CHEM Study, to name but a few) stress inquiry teaching methods as a main teaching strategy for learning. Yet traditional didactic teaching continues to predominate in classrooms throughout the country. Why does this condition still exist in light of what has been reported above? This investigation sought to shed some light on this stated condition and provide a possible answer as to which is more effective: inquiry or didactic teaching.

Research Design and Procedure

A sample of 80 students, 42 females and 38 males, from the fourth and fifth grades of a small town elementary school were randomly assigned to four treatment groups. This controlled for both class and grade level. The students were divided into 16 groups. Ten groups had five pupils, three groups had six students, and three groups had four students. This condition of uneven numbers was due to absences and unequal numbers of pupils in the classrooms used in the study.

The experimental design was a 2 x 2 factorial design in which learning structure (cooperative and competititve) was orthogonally crossed with teaching style (didactic and inquiry) which results in four treatment groups.

The inquiry teachers were trained in the inquiry strategy: (a) to ask questions with several possible answers, (b) to remain quiet at least three seconds after asking a question, (c) to invite students to answer their own questions, (d) to encourage students to consider questions, and (e) to ask them to do their own summarizing and interpreting. Teachers trained in the didactic strategy were asked to: (a) ask questions that encourage a single response, (b) evaluate the correctness of students' response, (c) demonstrate and explain information, (d) volunteer unsolicited information, and (e) answer questions authoritatively.

The investigators measured the dependent variables of acceptance of teaching method, acceptance of teacher, and perception of peer support on a three-point-self-report scale. Students also reported the extent to which they felt they had learned (subjective learning). All testing was done at the completion of a one hour lesson, the subject of which was a lesson on liquid evaporation adapted from Science for the Seventies. The test items were judged bv "relevant experts to have content validity and was found to have a test-retest reliability coefficient of 0.84."

Each group of students was escorted from their own classroom to another classroom in the school building. They were taught by one of three female undergraduate students who were specially trained in didactic and inquiry methods. Each lesson lasted about 55 minutes after the investigators left the room. At the completion of the lesson the investigators returned to the classroom and administered the test instruments. Teachers were randomly assigned to groups and none were informed of the research hypotheses in advance of the study. No teacher taught any condition more than twice. After the testing students were requested not to discuss the lesson with other students. They were then returned to their regular classroom. A week after the sessions the investigators returned to the classroom to discuss the entire study and give each student who participated a small prize.

To insure that teachers did teach either the didactic or inquiry method when required, a script was designed. Guidelines for both inquiry and didactic teaching were followed in constructing the scripts. The scripts were reviewed by two science educators familiar with both teaching methods and revised to their satisfaction.

Findings

To test the four hypotheses under the conditions described above, testing occurred at the conclusion of the lessons. Results of the analyses of data collected on the four dependent variables revealed no significant differences between fourth and fifth graders. Data were then collected for every grade for additional analyses.

Using a 2 x 2 ANOVA on student acceptance of teaching method revealed a significant main effect for teaching style, a significant main effect for learning structure, and a significant interaction effect, all at the 0.01 level. Thus acceptance of teaching method is a function of the interaction of learning structure and learning style. Follow-up tests revealed that students in the competitive-inquiry classroom disapproved of the way the lesson was taught significantly more than did students in the cooperative-inquiry condition, students in the competitive-didactic condition, and students in the cooperative-didactic condition, at the 0.01 level of significance.

Analyses of student approval of the teacher was not significant at the 0.05 level of significance (actually $p < 0.08$). Thus approval of the teacher is not dependent upon student acceptance of the teaching method.

It was also revealed that students in the cooperative condition liked being with other students in the session significantly more than did students in the competitive condition. In addition, cooperative students believed they learned significantly more in the sessions than did students who were in the competitive condition. And, finally, students in the competitive condition did not rate their acceptance of the didactic teaching method significantly greater than students in the cooperative condition.

Interpretations

From the results the investigators suggested that the traditional competition in the American classroom is not compatible with inquiry teaching methods. When compared with the other three conditions, students competing with each other in an inquiry situation tended to reject the teaching method. Students also disapproved of the teacher. Inquiry on the part of students requires peer support while competition interferes with the process, and thus cooperation may be a requisite for inquiry learning.

A major hypothesis of the study was that inquiry teaching was more acceptable under cooperative than under competitive conditions and didactic teaching, more effective under competitive conditions. While the interaction of the conditions was significant at the 0.01 level, the difference in didactic conditions was not significant. In other words, inquiry teaching was more acceptable under cooperative conditions, and neither cooperative or competitive conditions significantly affected student acceptance of didactic teaching methods. The same was true for teacher acceptance. It has been shown that cooperative conditions tend to foster positive attitude to both the teaching method and the teacher. This condition results in students perceiving didactic or inquiry teaching methods positively.

The investigators note that methods in the study may not have adequately tested both didactic conditions. The reason being that interesting material was presented for a short period of time by the didactic teachers. And if the same teachers taught less interesting material for a longer period, different results might have been obtained.

Cooperation does promote positive attitude toward teaching methods and teacher, but competition in this study does not uniformly promote negative results. Students in the competitive condition accepted the didactic method more than the inquiry method.

ABSTRACTOR'S ANALYSIS

Another article that belongs to an ever-expanding number of inquiry studies. The article does address a basic concern in science education: conditions that may facilitate science inquiry teaching. While it does not provide an answer to the main question raised by the authors, it does suggest conditions that may be required for science inquiry teaching. Several questions do arise, however, as one reads through the article. What are the research hypotheses? Do the results support the hypotheses? The answers to these questions are not easily found.

It appears that the authors have combined the research hypotheses, review of the research, purpose of the study, and significance of the study. Why didn't the authors place this information in separate sections instead of combining them? This abstractor is aware of the space limitations placed on writers when submitting materials for publication in a research journal. This may well explain why the discussion section also appears to be somewhat discombobulated. Adequate space in journals is required for a thorough and adequate reporting of research results. This is a problem that confronts all writers and researchers in these days of cost consciousness.

There are additional questions raised besides those listed above. How were the school, classes, and sample chosen? How were the teachers for the investigation chosen? How were they trained and for how long? What were the characteristics of the population? Of the sample? The reader is entitled to know this information. Understanding is enhanced when these data are recorded in the research design section. Questions such as these are critical to other researchers. This information is important when evaluating the results. Generalizability may be enhanced when this has been included in the research report. Thus a major criticism of this study lies in an inadequate reporting of the research design in which the authors do not describe their sampling procedures nor describe the student population used.

While authors are not responsible for the variability that exists for population and sample descriptions in journals, editors must take the responsibility. They must set standards that can be used by writers and researchers preparing research reports for journal publication. Because this critical information is required for research replication, these standards or guidelines must be forthcoming.

Several studies of a similar nature are identified in the review of the literature at the beginning of the article. They might have been discussed in greater detail in order to place the current study in proper prospective. But this is not done adequately. And when the results are discussed, little is done to relate the results to similar research. Now this is important when attempting to understand this study in relation to the matrix of similar research. This is an important part of preparing research reports since the reader must determine their significance and relationship to prior research.

In addition to some of the points raised above, others are brought to the attention of this reader in the discussion section. The results are indeed limited by the study's sample and the conditions described and defined. But this is the case with much research. In addition, other factors that may have biased the results were the "teachers," the duration of the study, and the materials used in the study. Some or all of these factors must be examined in future research if the accuracy or findings of the study are to be accepted.

In summary, the study does examine an interesting question about the conditions necessary for the acceptance of inquiry teaching by students. It also raises many additional questions that can only be answered through extended research efforts. In the space permitted, the authors do report the results of their study. With the exception of the questions raised and the few research design flaws identified, the study does point the way for additional research needed to understand the conditions required for science inquiry teaching. But it is doubtful that this line of research will ever provide an adequate answer to Dewey's question about teaching methods currently used in schools.

Ryman, Don. "Teaching Methods, Intelligence, and Gender Factors in Pupil Achievement on a Classification Task." Journal of Research in Science Teaching 14(5): 401-409, 1977.

Descriptors--*Cognitive Development; *Educational Research; *Intelligence; Junior High School Students; Science Education; Secondary Education; *Secondary School Science; *Sex Differences; Teaching Methods

Expanded abstract and analysis prepared especially for I.S.E. by Russell H. Yeany, University of Georgia.

Purpose

The study was conducted to investigate variations in teaching method, level of intelligence, and gender as predictors of pupil achievement as measured through a task involving a hierarchical classification of organisms.

Rationale

The educational setting of the study was in England where grammar and secondary/modern schools were being reorganized into a comprehensive system. This reorganization led to more mixed ability and gender groups. But many teachers continued with a single traditional teaching approach. At the same time, the Nuffield biology program was being adopted in some systems. This program recommended an inductive discovery-based teaching strategy which the author stated should theoretically lead to improved classification abilities.

Also, the author cited evidence of interactions between the method of instruction and pupil intelligence and some gender difference as measured in problem-solving tasks.

Research Design and Procedure

The sample was drawn from four comprehensive secondary schools as mentioned above. Four of the schools had adopted the Nuffield approach.

Two of these were selected at random for comparison against two non-Nuffield schools which were also randomly selected. Two first-year biology classes were then randomly selected from each of the four schools and subjects were stratified according to gender and intelligence and randomly chosen to provide a total of 96 pupils in two teaching methods, two intelligence groups and two genders. The average age of time of testing was 12 years and 5 months.

Stratification of the intelligence groups was made on the basis of a group test with a split for above and below mean intelligence.

The classification task involved 38 drawings of living organisms similar to those in the Nuffield text. The pupils were asked to categorize the organisms into all the groups in which they belonged. In total, 89 codings of the 38 organisms was possible. A pre-test was conducted to ensure that the pupils were generally familiar with the specific names of the organisms depicted by the drawings.

Findings

The study reported a second-order interaction among methods, intelligence and gender, a first-order interaction between methods and intelligence and a significant main effect which favored the above average intelligence group as measured by classification abilities.

Interpretations

In relation to the second-order interaction, the author concluded that the use of Nuffield methods seems unsuitable for girls of below-average ability and that girls of above-average ability seem to under-achieve when taught by traditional methods. The existence of the interaction restricted the interpretation of the main effect and the first-order interaction.

ABSTRACTOR'S ANALYSIS

This research can be classified as an ATI (Aptitude Treatment Interaction) study and is representative of an important set of questions which need to be addressed in science education research. That is, what type of instruction is effective for which types of pupils? Not simply what type of instruction influences the mean of the group?

The report was generally well written and the findings did indicate that there probably are some differential effects of teaching methods across the various intelligence levels and genders of the pupils. Unfortunately, there was no adequate operational definition of the variations in teaching methods nor any evidence that the methods of instruction were monitored to determine what was happening in the classrooms in terms of teacher/pupil behaviors. This is a common flaw in many studies and needs to be more regularly attended to.

The author concluded that his findings support the view that pupils should be taught biology in homogeneous groups in terms of intelligence levels and gender. But he also recognized that most biology classes are co-educational and of mixed-ability levels and suggests that teachers should select their methods with care. However, there is nothing clear in the report that would indicate the specific nature of the methods that should be selected for various student types.

A possible selection bias exists because the teaching method was not randomly assigned to schools or subjects. Two schools were randomly selected from four schools that had adopted the Nuffield program prior to the study. It is very possible that basic differences existed between the Nuffield and traditional schools that might have influenced the data. This particular limitation of the study is faced by many educational researchers. It does not mean that the studies should not be run, but we should be cautious about our interpretation.

Also, it is becoming less acceptable to analyze continuous data as a dichotomy when conducting ATI research. In this study, considerable precision in the data was lost when the intelligence data were stratified into above and below average groups for the analysis. The probability of misclassification of pupils who scored near the mean on the intelligence test is great.

The Ryman study pointed out some possible conditions under which teaching methods have differential effects. The findings are not clear enough to guide extensive revisions in classroom practice but it should serve as a basis for further research.

Future studies in science education which examine aptitude by treatment interactions can provide much information regarding the selection of teaching strategies and matching them to the situations where the greatest amount of learning can occur. These studies should include very explicit definitions of the strategies being employed so that the practitioner can better apply the findings to the classroom. When possible, the pupil aptitude variables should be analyzed as continuous variables (regression analysis using general linear models can facilitate this) so that precision in the data is not lost and the classification of pupils as "highs" or "lows" is not so arbitrary or capricious.

Romero, Frank S. "The Effects of Auto-Tutorial Science Process Instructions on Teacher Achievement and Its Relation to Specific Undergraduate Majors." Journal of Research in Science Teaching, 14(4): 305-309, 1977.

Descriptors--*Achievement; *Autoinstructional Methods; *Education Majors; Elementary School Science; Higher Education; *Instruction; Lecture; *Methods Courses; *Preservice Education; Science Education; Teaching Methods

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

Purpose

The stated purpose of the research report was to evaluate the effects of two methods of instruction, lecture approach vs. auto-tutorial approach, on the achievement of preservice teachers in five processes of science representing three different majors. The undergraduate majors were science, humanities, and social studies.

Rationale

While there have been conflicting research reports concerning the value of auto-tutorial instruction, some research suggests that the auto-tutorial approach for preparing teachers to use the science process skills may be promising. A few studies (three) conclude that the method has been used to develop teachers competent in the use of science process skills in their classroom teaching. However, no empirical research exists to indicate that the auto-tutorial method has been used with undergraduates of different undergraduate majors.

Research Design and Procedure

For the purpose of this study the author used the posttest-only control group diagrammed below:

R	X	0
R		0

The study included 54 undergraduates equally divided on the basis of undergraduate major (i.e., 18 with a science major, 18 with a social studies major, and 18 with a humanities major). They were randomly assigned to the experimental (auto-tutorial) and control (lecture) groups by a 2:1 ratio respectively. Instructors for the course, each with at least five years experience in the teacher preparation program, were randomly assigned to the two groups.

Both groups participated in the study for one month. Each group met three times per week during the study and, at the conclusion of the study, the measurement instruments were administered to both groups.

The overall GPA and GRE scores were identified and summarized for close scrutiny. No apparent statistical analyses were reported to have been done on the descriptive statistics. It can be inferred that the scores were reasonably similar and that the groups were homogeneous except for major.

Instructional materials for teaching the five science processes (i.e., observing objects, reporting data in an organized form, organizing objects with a variety of attributes, experimenting and testing hypotheses, and inferring and generalizing from empirical data) were developed for the students to use. The scientific method model was employed throughout the activities. The materials were based on those used in the new elementary science curricula and developed in corroboration with the science departments at the university in which the study was conducted.

An instrument (previously developed but modified for this study) to measure the competence of students to use the processes of science was employed and had a reported reliability coefficient of 0.89. Content validity was determined using factor analysis. However, the results were not reported in the study. The instrument was modified by altering the scoring system and by adding 10 activities congruent with elementary science curricula inquiry activities. The activities were developed by graduate students and then judged by five researchers to determine if the activities were valid with respect to the content. The activities were randomly chosen to construct five minitests to measure each of the five process skills. Reliability coefficients were 0.92 (observing), 0.90 (reporting), 0.90 (organizing), 0.91 (experimenting), and 0.93 (inferring).

The tests were randomly ordered for each student and administered at the conclusion of the instruction period. The instructor, the curriculum developer, and the supervisor of student teachers scored each test using a scale of 0-3. The scores were averaged for each student's final score. No interrater reliability coefficients were determined for the scorers.

Findings

The investigator employed multivariate analyses of variance (MANOVA) to analyze the data. There was a significant difference between the auto-tutorial (experimental) and lecture (control) groups in favor of the auto-tutorial method. There were significant differences between method and undergraduate major in favor of the science undergraduates. There was a significant difference between the auto-tutorial science majors and the auto-tutorial humanities and social studies groups. The science group was significantly higher in overall achievement in the five tests as compared to the other two groups. There were no significant differences between the latter two groups. While the science undergraduates improved significantly as a result of the auto-tutorial treatment, the other two groups also improved significantly as a result of exposure to the auto-tutorial method. But the gain was not statistically significant. Thus the auto-tutorial method is an effective method for improving students' achievement of the five process skills identified in this study.

Interpretations

The research demonstrates that the auto-tutorial approach to teaching science process skills is an effective method when the students are either science or nonscience undergraduate majors. This is so when compared to the lecture method used with similar students representing the same undergraduate majors. Based on these results, the auto-tutorial method should be used to instruct preservice teachers in the science inquiry skills. The investigator concludes by suggesting that there is no alternative.

ABSTRACTOR'S ANALYSIS

The report just described and summarized above is neat, short, and straightforward. However, a few questions do arise after reading it. For instance, the investigator never really identifies the problem that is to be investigated. Three studies are identi′ ed but no information is provided as to what exactly happens in these studies. So the justification for the study is strained a little.

The investigator fails to state how the students were chosen. It is pointed out early that they are randomly distributed into the two groups (control and experimental). So the reader is left wondering how the subjects for the study were chosen. No reason is given for the choice of group characteristics reported in the written report. Why didn't the investigator identify the number of years in attendance (junior, senior, other), or age of the subjects? These factors are important especially if replication of the study is to be undertaken.

It would have been helpful to review in greater detail the materials developed and used in the study. It is important to provide sufficient information so that the reader knows and understands what is being done or attempted with the experimental group. It would be necessary here to write to the investigator and obtain the materials used if the study was to be replicated.

While the investigator does state the length of the study, it is unclear as to how long the students meet during each of the four weeks required for the treatment. It is not quite clear as to the environment in which the study is conducted. For instance, it is conducted in a standard college classroom, lecture hall, or methods or science laboratory? This is important due to the nature of the treatment.

The instruments reported for use in the study have been modified. But it is not quite clear how these changes are made and why. The reliability coefficients for the modified instrument are very good and should provide the information required by the investigator. But the scoring system is not fully explained and is really left up to the imagination of the reader. Further, the tests were marked by three different people and then the results were simply averaged for each subject participating in the study. This does average out some error but is not a good way for determining the accuracy of the scoring system. It would have been better to make a random sample of the tests, score them, and then determine an interrater reliability coefficient. This would have been better than the method employed and a little more economical in terms of time requirements.

The results of the study are presented in one simple chart. However, in using MANOVA the investigator does not report if there are any interactions. MANOVA is a statistical method designed to uncover interactions. But these are never mentioned. It would have been good to have presented a few more charts summarizing the results. The article could be improved in this area.

In discussing the results, the investigator draws a reasonable inference concerning the results reported. However, there is never any mention as to the effect undergraduate major has on achievement in the use of science process skills. Surely the three studies identified in the beginning of this analysis had subjects who had similar or identical undergraduate majors. But no connection is made here concerning this variable. The investigator does state that research has "failed to produce any empirical evidence that the auto-tutorial approach was successful in increasing achievement of student teachers representing undergraduate majors" (p. 305). This really means that no meaningful problem exists and none was ever identified in the article. The investigator must be careful to see that the problem is precisely stated and identified.

This abstractor recognizes the space limitations that are placed upon investigators when submitting manuscripts for inclusion in professional journals. But sufficient space must be allowed to present reasonable detail and information required for proper communication to professionals in the field as well as for replication purposes. Indeed, the items identified above in this analysis do not undermine the research conducted here. Some of the questions raised above, when answered, may provide responses designed to improve the preciseness of research reports submitted for inclusion in professional journals.

Santiesteban, A. and J. Koran. "Acquisition of Science Teaching Skills Through Psychological Modeling and Concomitant Student Learning." Journal of Research in Science Teaching, 14(3): 199-207, 1979.

Descriptors --`c ievement; Audiovisual Programs; *Educational Research; *Ins. tion; *Performance Based Teacher Education; Preservice Education; *Role Models; Science Education; Science Teachers; *Teacher Education

Expanded abstract and analysis prepared especially for I.S.E. by David F. Butts, The University of Georgia.

Purpose

This study had a dual purpose:

1) to compare the effects of videotape and audio-mediated models on the acquisition of teaching skills;
2) to validate the teaching skills in terms of student learning.

Rationale

Modeling is one way to influence behavior -- including those teaching behaviors of science teachers. Modeling may be by examples, or by video, written or audio representations. A second related rationale depicts the need to validate acquired teaching behavior in terms of changes they make in student outcomes.

Research Design and Procedure

With a sampling of 48 preservice teachers and 184 third and fourth grade students, a post-test-only control group design with random assignment of teachers and students to two treatments (video modeling and audio modeling) and one no-treatment group was used. A student-with-no-instruction group was used to compare achievement of student outcomes. The outcome measures were of teacher knowledge and student knowledge.

X_{vm}	$O_{1,2}$
X_{am}	$O_{2,4}$
X_c	$O_{5,6}$
	O_7

Where X_{vm} = Video model treatment

X_{am} = Audio model treatment

C = No model treatment

$O_{1,3,5}$ = Measure of teacher criterion test

$O_{2,4,6,7}$ = Measure of student process test.

In each treatment group, an introduction was made to the task of using observing and classifying questions. In the two treatment groups a model was then presented. Following thi -tep, all three groups developed a lesson and taught it in a microteaching session. A post-test was then given to both teachers and students.

Findings

1. Based on the teacher criterion test, both modeling treatments were more effective than the no-model treatment.
2. Audio model teaching also scored better on the teacher criterion test.
3. The modeling treatment were superior to the control group as reflected in the students' performance.

Interpretation

Both modeling treatments were equally effective in helping preservice teachers acquire specific teaching behavior. While the video model had the potential of a richer source of input for the teachers than the audio model, that richness appears not to be related to the teacher's acquisition of behavior. For verbal teaching behaviors, the audio model appears quite adequate. Some evidence exists that students pick up the questioning behaviors modeled by their teachers.

ABSTRACTOR'S ANALYSIS

The first purpose of this study is carefully articulated with referenced studies in the area of modeling studies. The second purpose of the study could have had a more substantial theoretical basis if it was really an important aspect of the research. That teachers model the behavior that is presented them is confirmed. That students may indeed model the teacher's verbal behavior is an exciting fresh dimension in this study.

The design of the study enables the reader to have high confidence that the manipulated variables were indeed different, that the teacher behavior was indeed both acquired and used in the microteaching setting, and that students also demonstrated the behavior. A pretest would have given more confidence to the reader that the teacher and student behaviors were the result of the treatment and not of previous experience.

The validity of the measures of the dependent variables is missing so the reader must use caution in interpreting the conclusions.

The report of this useful and important study is clear. The authors have communicated well in sharing both the questions and the research results that fit together in their answer to the questions.

Canary, Pat; Carol Hudachek; and Robert D. Allen. "Student Response to Extra Credit Opportunities in a General Biology Course." Journal of College Science Teaching, 4(4): 312-314, 1976.

Descriptors--Achievement; College Science; Educational Research; Higher Education; *Instruction; *Motivation; Science Education; *Student Motivation; *Teaching Methods

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Jacqueline Sherris and Jane Butler Kahle, Purdue University.

Purpose

The purpose of this study was to investigate the utilization of extra credit opportunities by college freshmen in a large enrollment biology course.

Rationale

The impetus for this study was a desire to learn more about the relationship between extra credit work and motivation in a general biology course. Although McKeachie (1969) concluded that grades were the primary motivational device available to teachers, the authors suggested that grades were more threatening than motivating. Furthermore, they cited the limited bonus system utilized by Postlethwait et al. (1972) and Hurst and Postlethwait (1971) as an example of extra credit work to improve grades in large biology courses. Although the results of these studies did not fully support the assumption that extra credit work had motivational value, the authors emphasized the need for more information regarding the use of extra credit and its potential for increasing student motivation.

Research Design and Procedure

This was a descriptive study in which the relationships between student achievement and use of and/or success on extra credit units were observed. No comparison group was present. The variables investigated for each student included the following ones:

1. Operational course grade: point total for the course from exam and laboratory scores excluding extra credit points.

2. Number of extra credit units attempted.

3. Number of extra credit units successfully completed.

4. Composite ACT scores.

The subjects of the study were freshman college students enrolled in a general biology course. The sample consisted of both majors and nonmajors. The total number of subjects was not stated but, from a table illustrating the number of students in each operational grade category, we assumed that N = 1231.

The one-semester general biology course in this study required students to attend three video-taped lectures and one two-hour laboratory each week. The course guide was based on laboratory and examination scores equaling a total of 400 points. The eight extra credit units consisted of five Scientific American articles, two two-hour audio-tutorial minicourses, and one film. The film and minicourses were evaluated by a posttest consisting of 10 multiple-choice items, and the Scientific American articles were evaluated by a posttest consisting of two essay questions. If students scored at least 80 percent on the extra credit posttests, they received five bonus points per unit. If not, a zero was recorded for that extra credit unit.

The authors summarized the data on a chart using the following headings: operational grade, number of students, percent attempting one or more extra credit units, percent of those attempting one or more extra credit units who are successful. Then, the latter two categories were plotted against operational grade. The data were further analyzed by examining the linear regression between operational point total and both the number of extra credit units attempted and the number of extra credit units successfully completed. The latter two categories also were examined separately in relation to operational point total. It is not clear whether simple correlations or regression equations were calculated for these relationships. A further regression equation predicting operational point total was calculated using 626 students for the latter two variables listed above and composite ACT scores. There is no indication how or why this subgroup was selected.

A final observation was the amount of time spent by staff in the preparation, administration and evaluation of extra credit units.

Findings

The authors' findings were the following:

1. More students achieving operational grades of "B" attempted extra credit units (62.6 percent) than did students achieving operational grades of "A," "C," "D," or "F."

2. More students achieving operational grades of "A" were successful in completing attempted extra credit units (92.3 percent) than were students achieving operational grades of "B," "C," "D," or "F."

3. There was a highly significant relationship between operational grade and both number of extra credit units attempted and number of extra credit units successfully completed ($p < .001$). There were also highly significant relationships between operational grade and each of the latter variables separately (in both cases $p < .0001$).

4. A total of 41.5 percent of the variance in the operational point totals was explained by a multiple regression equation using three independent variables and operational point total as the dependent variable. Composite ACT scores were the best predictor, number of extra credit units completed the next best predictor, and number of extra credit units attempted the least useful predictor.

5. Staff time spent in developing, administering and evaluating extra credit units was about 50 hours per minicourse and 14 hours per Scientific American unit. No analysis was made for the film.

Interpretations

The authors concluded that the large number of students who attempted extra credit units indicated considerable motivation resulting from the extra credit offerings. They also suggested that students achieving lower grades exhibited less motivation. Also, it was concluded that successful students work harder than less successful students and thus were more likely to try extra credit work.

The authors suggest that the use of extra credit materials in college courses involves a consideration of the following factors: staff time involved in handling these materials, student attitudes concerning extra credit materials, and motivation effects on successful and unsuccessful students.

ABSTRACTOR'S ANALYSIS

In their introduction, the authors stated a desire to learn more about factors affecting student motivation. Since it has been established that student motivation may be a major contributor to educational success, continued research in this area is appropriate (Hubbard, 1974; Hunt and Hardt, 1969). The initial goal of the study, to look at the relationship between motivation and extra credit work, showed promise. However, the study's design and experimental procedures resulted in serious distortion of the initial goal and, consequently, in difficulty in drawing clear conclusions from the data obtained.

The present reviewers could not identify the basic question asked and, therefore, could not determine a rationale for the experimental procedures carried out in this study. First, although the authors were interested in investigating extra credit and its relationship to motivation, the reviewers were unable to draw conclusions about this hypothesized relationship since no attempt was made to measure individual motivational levels. Although it was suggested that grades may be an indicator of motivational level, this relationship was not clarified. Furthermore, no attitudinal data were collected from students.

A second possible question was suggested by the effort to establish the best predictor of total course grade. A substantial part of the statistical analysis used multiple regression techniques to attempt to isolate factors important in predicting total points obtained in the course. The factors included in the regression equations were extra credit units attempted, extra credit units successfully completed, and ACT scores. The relevance of these regression analyses to student motivation was not elaborated.

A third area which was briefly mentioned by the authors was the time and/or cost effectiveness of extra credit work, particularly in regard to staff time devoted to the preparation and evaluation of the units. This question was referred to in the results section for two of the three types of extra credit units offered. However, the authors did not draw any conclusions concerning the management aspects of extra credit work.

This investigation was a nonexperimental, descriptive study. No comparison group was present, and thus the authors primarily were able to study the data through descriptive and correlational statistics. There was no dependent variable isolated, although total course points functioned as a dependent variable in much of the statistical analysis. The major variable of interest was the extra credit units. Eight different units were treated as equivalent, but insufficient information was given to establish equivalency. It might have been more appropriate to determine if the three types of extra credit activities (films, Scientific American articles, and minicourses) and the associated evaluation instruments (essay or multiple-choice tests) were approached equally by the students.

The data were presented in chart and graph form with both forms illustrating that better students (in terms of point total) successfully completed more extra credit units than did less able students. Based on the data this conclusion is correct, but these reviewers think that the percentage shown by the authors hides much of the effort put forth by less successful students in attempting and in completing extra credit units. We have added to the authors' data table the following categories:

one, actual numbers of students and, two, percent of total students succeeding in at least one extra credit unit (in view of a percent of students in each grade category). These revisions were added to the authors' chart entitled "The Percent of Students in Each Operational Grade Category Attempting and the Percent Succeeding in Extra Credit Work." The revised table presents more information, derived from the initial tabulation of data.

TABLE 1

Operational Grade	Number of Students	Number Attempting Extra Credit* Percent Attempting)	Number Succeeding in Extra Credit	Percent Attempting Who Succeed in Extra Credit*	% Total Succeeding (456) in Each Category
A	120	65 (54%)	60	92%	13.15%
B	318	199 (63%)	179	90%	39.25%
C	474	208 (44%)	160	77%	35.10%
D	252	76 (30%)	50	66%	11.00%
F	67	13 (20%)	7	65%	1.50%
	1231	561	456		

*Indicates at least one extra credit unit.

As seen in the last column of Table 1, by far the largest percentage of total students who succeeded in at least one extra credit unit were "B" and "C" students (74.35 percent). The "A" students made up 13.15 percent of the total, and constituted only a slightly larger percentage than did the "D" students. If we assume that most college students who achieve "F" grades are affected by many extraneous factors and thus reasonably may be discounted in a descriptive study such as this, it is obvious that the bulk of students who took advantage of extra credit opportunities were students receiving "B," "C," or "D" grades. Furthermore, one may presume that these were the students whom the authors were especially interested in motivating. The authors established that an individual "A" student was more likely to successfully complete an extra credit unit than an individual student in any other grade category. But if we look at the class as a whole, we see that a large number of "non-A" students were motivated enough to attempt and succeed in at least one extra credit unit. These observations do not contradict the authors' conclusions, rather they offer information from which to draw additional conclusions.

The regression analyses of the data were appropriate in the examination of some factors which predict total exam score, in particular the prediction value of the number of extra credit attempts and the number of extra credit successes. The description of the regression analyses was brief and could have been augmented by a table. These reviewers found parts of this discussion confusing. For example, information explaining the use of only 626 of the 1231 sample in the multiple regression analysis was not included.

The authors may have more data from this study than were utilized in this brief article. Additional information about student use of extra credit units could have allowed for more conclusions concerning the effect of extra credit work on student motivational levels. The mean number of units attempted and the mean number of units successfully completed at each grade level would have been informative. The regression analyses indicated that the number of units attempted and the number of units successfully completed were highly correlated to total course points, but one does not know how much of the total variance is accounted for by each of these factors. The reader can only guess as to whether a student who attempted extra credit units attempted one unit or eight units. This kind of information possibly could reveal more about the motivational effects of extra credit work. The reader also does not know what proportion of students who successfully completed extra credit units also improved their letter grade by the extra credit work. This information might have enabled the authors to make more solid conclusions about the relationship between student motivation and grades.

The results of this study allowed the authors to make some conclusions about the use of extra credit units by students in different grade categories. Few conclusions were made concerning the effect of extra credit opportunities on student motivational levels, the expressed goal of the study. Since the authors still may have access to the large general biology class used in the study, it would be informative to design an experimental or quasi-experimental study investigating the effect of extra credit units on student motivation. Student motivational level would need to be specifically measured by recording the numbers of extra credit units attempted and successfully completed for each student, and by collecting attitudinal data concerning the extra credit units, and the course as a whole. The general biology class undoubtedly was divided in some manner, perhaps by quiz sections or lab groups, and it would be fairly routine to design a true or quasi-experiment offering extra credit units to one group and not to another one. If this procedure were perceived as unfair by the students, a time series quasi-experimental design could be employed. With this design, all students would receive the same treatment and every other unit in the course would offer extra credit units. Thus student attitude and achievement when working on units without extra credit offerings could be compared to student attitude and achievement when working on units with extra credit offerings. Although these designs have validity problems, more could be learned about a specific factor (i.e., motivational level) from them than from a noncomparison group study.

REFERENCES

Hubbard, L. R. "The Effects of Audio-Tutorial Introductory Biology on the Attitudes and Self-concepts of Prospective Elementary Teachers." Unpublished doctoral dissertation, Purdue University, 1974.

Hunt, D. E. and R. H. Hardt. "The Effect of Upward Bound Programs on the Attitudes, Motivation, and Academic Achievement of Negro Students." Journal of Social Issues, 25(5): 117-129, 1969.

Hurst, R. N. and S. N. Postlethwait. "Minicourses at Purdue: An Interim Report," in The Use of Modules in College Biology Teaching, J. G. Creager and D. L. Murray (Eds.). The Commission on Undergraduate Education in the Biological Sciences, 1971.

McKeachie, W. J. Teaching Tips: A Guidebook for the Beginning College Teacher, 6th ed. Lexington, Massachusetts: D. C. Heath and Co., 1969.

Postlethwait, S. N.; J. Novak; and H. T. Murray. The Audio-Tutorial Approach to Learning, 3rd ed. Minneapolis: Burgess Publishing Co., 1972.

Ben-Zvi, R.; A. Hofstein; and D. Samuel. "The Effectiveness of Filmed Experiments in High School Chemical Education." Journal of Chemical Education, 53(8): 518-520, August 1976.

Descriptors--*Chemistry; Educational Research; Films; *Instruction; *Instructional Films; Science Education; *Secondary Education; *Secondary School Science

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Chris Pouler, Hyattsville, Maryland.

Purpose

This study compared the use of filmed experiments as an alternative to study-centered laboratory work. Specifically, the researchers determined the effectiveness of these filmed experiments to actual laboratory experiments. The observable student behaviors included:

1. General achievement in chemistry (knowledge and understanding of subject matter).

2. Knowledge of principles underlying chemical experiments and laboratory techniques.

3. Manipulative skills relating to the handling of equipment and use of apparatus.

4. Observational attainment and problem-solving abilities in relation to laboratory situations.

Rationale

Laboratory work is considered a valuable tool for students to discover the facts and concepts of science. In fact, many of the contemporary curricula emphasize laboratory experiments as an integral part of the learning process. Because all schools may not be equipped to provide laboratory experiences the researchers studied the extent to which filmed experiments could replace actual student experimentation. The research could then be applied to support the use of filmed experiments as a viable substitute for laboratory work.

Research Design and Procedure

Population. The sample involved 330 tenth-grade chemistry students from six different high schools in Israel. The 130 males and 200 females were divided into two groups--film and experimental. The film group contained a total of 150 students and the experimental group, 180 students. Pretests measuring (a) IQ, (b) Science Interest, and (c) Science Attitudes were administered. The results were used to establish equivalence of the two groups. For five months, both groups were exposed to identical learning experiences except that one group did laboratory experiments while the other observed filmed experiments. Eleven experiments were covered which related to the concepts of

(a) mass and volume relationships, (b) oxidation and reduction, and (c) atomic structure. The filmed experiments consisted of four-minute silent film loops that portrayed the experiments in a similar manner as the actual laboratory investigations. The film loops were silent, color presentations with Hebrew captions. Follow-up activities for the film group and the experimental group were identical. At the conclusion of the five-month investigation period, all participants were post-tested.

Instruments. The researcher developed five different instruments to assess the effect of the treatment.

1. Achievement in Chemistry Test. This 25-item test included the general information taught during the investigatory period. The test had an average difficulty index of 0.61 and a Kuder-Richardson reliability of 0.78.

2. Specific Knowledge Test. This test was divided into two sections (Principles and Techniques/Methodology) and was intended to assess students' knowledge and understanding of experimental techniques and work.

3. Practical Test 1. This test was intended to determine the students' manipulative skills. Specifically four areas were measured--(1) Experimental technique, (b) procedure, (c) manual dexterity, and (d) orderliness.

4. Practical Test 2. An exercise involving the quantitative investigation of the effects of heat on cadmium carbonate was developed to examine students' skills in dealing with a practical problem-solving situation. Students were expected to plan and conduct the experiment.

5. Observation Test. Students had to watch six test-tube reactions and write as many observations as possible. There were 23 possible reactions, 13 of which dealt with color changes. Separate scores were computed for color and noncolor changes.

Findings

Between the film group and the experimental group there were no significant differences for (a) Achievement in Chemistry, (b) Specific Knowledge of Principles, (c) Specific Knowledge of Technique/Methods, (d) Practical test--Problem Solving, and (e) the Observation Test. Because significant differences did occur on the Practical Tests--Manipulation, it is worthwhile to describe each aspect of this test.

1. Experimental Technique involved the handling of apparatus and chemicals; safe execution of an experimental procedure; taking of adequate precautions to ensure reliable observations and results.

2. Procedure involved the correct sequencing of tasks forming part of an overall operation; effective and purposeful utilization of

equipment; efficient use of working time; ability to develop an acceptable working procedure on the basis of limited instructions.

3. Manual Dexterity involved swift and confident manner of execution of practical tasks; successful completion of an operation or its constituent parts.

4. Orderliness involved tidiness of the working area; good utilization of available bench-space: purposeful placing of apparatus and equipment (p. 519).

For the Experimental Techique determination, there was significance at the 0.005 level favoring the experimental group. Similarly for the Procedure test there was a 0.005 significant difference favoring the experimental group. The test for Orderliness also favored the experimental group with a significant difference of 0.01. However, there was no significant difference on the Manual Dexterity measure. Taken as a group of tests, the total significant difference of 0.001 favored the experimental group.

A few words regarding the Observation Tests are also in order. As reported the Observation Test consisted of two parts--(a) color changes and (b) noncolor changes. The film group performed better on the color changes to the significance level of 0.01. The experimental group performed better on the noncolor changes to the 0.05 level of significance. When both tests are averaged and included as one test, there is no significant difference. Future researchers should be warned, however, that the mean results for the Observation Tests appear to be reversed for the two groups. The printing error does not alter the lack of significance.

In summary, only on tests of students' manipulative abilities did the experimental group outperform the film group to levels of significance.

Interpretations

The research indicates that, except for the display of manipulative skills, students who watch film loops rather than performing experiments are not affected in an adverse manner. Specifically, these students achieve equally on cognitive or laboratory-based problem-solving achievement. The disadvantage of film loops is apparent in the area of manipulative skills. " But the relative advantage gained by experiment group students over filmed experiments is small and points strongly to the potential of filmed experiments as a means of teaching manipulative skills" (p. 520). The authors' conclusion that "well-designed films or film-loops are a viable alternative to student-based laboratory work" (p. 520) appears to be supported by the data.

ABSTRACTOR'S ANALYSIS

This research report was short yet complete in the presentation of the experiment and the results. The research techniques appear to be sound.

The design provided as random grouping as possible. Six different high schools provided diversity of population. However, all students had benefit of previous science instruction in physics and biology. Perhaps one reason for the results was this prior exposure to science. Also the themes selected for the film-loops (mass and volume relationships, oxidation and reduction, and atomic structure) might have some carryover from the previous instruction in science.

Therefore, the film loops and experiments may have been reinforcing previously learned materials and not presenting new information to be discovered. It would be interesting to pursue this study with different age groups and various levels of previous science instruction.

The instruments used to measure the teaching strategies seem to be locally developed. Since the comparisons involved students in the same schools, perhaps this is not a major matter. Further, the authors appear to have taken care to develop worthwhile tests. The explanations of the contents were quite clear. Future research must certainly take care to design tests that will be both valid and reliable. There is always the temptation in analyzing research of this type to question if the laboratory experience of film loop really presented information that could not be obtained from the textbook and classroom presentations. Of course, hands-on laboratory experience indicated a manipulative advantage for the experimental group. It must be noted, however, that the performance advantage was only 10 percent. Of course, educators must determine the importance of such an advantage. The prior exposure to biology or physics may have improved the students' laboratory experience even prior to the study. There was no effort to assess the pre-research manipulative abilities of the students. This would, of course, make such a study most cumbersome.

The basic purpose of the study was to verify if films could be used in lieu of laboratory experiments if laboratory facilities were limited. Of course the curriculum and teaching strategies must provide for discovery by both laboratory or film loop. Other interesting possibilities for future research exist: color versus noncolor films, video tapes versus films, sound versus captions, various forms of teaching to learn better by film loops than by experimentation. Perhaps for students who are not interested in laboratory experiments the film approach provides an alternative.

This study was most interesting. The results are useful as potential for future research.

CRITIQUES

and

RESPONSES

Rubba, Peter A. and Hans O. Andersen. "Development of an Instrument to Assess Secondary School Students' Understanding of the Nature of Scientific Knowledge." Science Education, 62(4): 449-458, 1978.

Descriptors--*Educational Research; *Instrumentation; Measurement; Research; *Science Education; *Scientific Literacy; Scientific Principles; Secondary Education; Secondary School Students; *Testing

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

Purpose

The purpose of this project was to develop, field test, and validate an instrument to assess secondary school students' understanding of the nature of scientific knowledge.

Rationale

The project is linked to the general acceptance of scientific literacy as a major goal of science education. Although several position statements are cited which have been offered as definitions of scientific literacy, the project is most closely tied to the delineation of dimensions of scientific literacy developed by Victor Showalter and colleagues through efforts at the Center for Unified Science Education (Showalter, 1974). The project is predicated on the fact that Showalter's definition of scientific literacy has been used as a basis for establishing program objectives in many of the nation's schools, but the definition has not been used often as a comprehensive guide to science instruction. Specifically, the investigators assert no reliable and valid instrument has been developed to assess science instruction with respect to Showalter's criteria. This project was limited to developing "an instrument to assess secondary school students' understanding of the nature of scientific knowledge--the first dimension of the Showalter definitions of scientific literacy."

Research Design and Procedures

The instrument was developed and field tested using a process with seven steps.

Step1: Establishing a Model of the Nature of Scientific Knowledge--Building upon Showalter's claim that nine identifiable factors underlie the nature of scientific knowledge, the investigators used a panel of three philosophers of science to develop and refine a more succinct factor structure. Their final model contained six factors and was labeled "A Model of the Nature of Scientific Knowledge." The factors were:

AMORAL	Scientific knowledge cannot be judged in a moral sense; only Man's application of scientific knowledge can.

CREATIVE	Scientific knowledge is a product of human intellect, the invention of which requires creative use of the scientific inquiry process.
DEVELOPMENTAL	Scientific knowledge is never "proven" but is in nature capable of change as more evidence is accumulated.
PARSIMONIOUS	Scientific knowledge tends toward simplicity with specific attempts being made to minimize the number of concepts necessary to explain the greatest possible number of observations.
TESTABLE	Scientific knowledge is available and amenable to public, empirical test.
UNIFIED	Scientific knowledge develops from an effort to understanding the ity of nature and is a systemized network of laws, theories, and concepts.

Step 2: Item Pool Preparation--Twelve to fourteen positive effect item statements and the same number of negative effect item statments were developed for each of the six factors using a Likert-scale format.

Step 3: First Item Refinement Reading Level--Working with nine sixth-grade students of comparable reading ability, the test statements were written at the junior high school reading level.

Step 4: Second Item Refinement: Form and Content--Items were refined for form and content using a panel of ten doctoral students in science education. Fifty-seven pairs of item statements remained at the end of this step.

Step 5: Third Item Refinement: A Tryout--Using a five-point Likert scale ranging from strongly agree to strongly disagree, the items were randomized and administered to 31 high-science-ability secondary juniors attending a summer institute. Some changes were made in the items as a result of feedback from the students.

Step 6: Item Selection Panel: Judged Item Content Validity--Content validity of the 114 items was judged against the six-factor Model by nine experts representing philosophers of science, science educators, scientists, secondary teachers, and psychometricians. The end result was 36 positive and 36 negative effect items (not necessarily item pairs) which were judged to measure respective factors in the Model.

Step 7: Field Testing and Item Selection--The 72 items were treated as in Step 5 and administered to 674 science students (general science, biology, chemistry, physics, and physiology) at a midwest high school. Forty-eight items were selected for the final instrument based upon calculations of the most discriminating and reliable combination of items. This instrument was named the "Nature of Scientific Knowledge Scale" (NSKS).

Findings (Instrument Characteristics)

Internal consistency estimates ranged from 0.65 when administered to 101 ninth-grade general science students to 0.89 when administered to 36 twelfth-grade advanced chemistry students. Test-retest reliability estimates ranged from 0.59 for 52 freshman general science students to 0.87 for 35 advanced chemistry secondary seniors.

Construct validity was examined using an ex post facto design. "Forty freshmen completing an introductory college philosophy of science course were expected to understand the nature of scientific knowledge better than 125 freshmen at the same university with no formal history and philosophy of science background who were completing a biology course for nonscience majors." Mean scores of the two groups on the NSKS and its subscales were compared using "a t-test technique for independent samples." On four of the six scales and overall, the two groups were found to be significantly different (Table I). The investigators accepted these findings as evidence of NSKS construct validity.

Table I

t-Test Comparison of NSKS Scores Between Biology and Philosophy of Science Group

Subscale/ Score	Biology (n = 125) x	S.D.	Philosophy of Science (n = 48) x	S.D.	t	p[a]
Amoral	26.38	4.14	26.55	5.41	0.20	0.838
Creative	25.89	4.70	24.85	6.43	1.11	0.271
Developmental	29.82	3.25	31.30	3.72	2.42	0.016*
Parsimonious	22.80	2.90	24.30	3.72	2.65	0.009**
Testable	30.44	3.65	31.80	3.67	2.05	0.042*
Unified	29.66	3.61	32.00	5.29	3.16	0.002**
NSKS	164.99	12.73	170.80	15.47	2.38	0.018*

a Two-tailed probability

*p 0.005

**p 0.01

ABSTRACTOR'S ANALYSIS

The authors' purpose--to develop an instrument to assess students' understandings of the nature of scientific knowledge--was predicated on their belief that no comparable instrument existed. This reviewer would agree that no specific and valid instrument did exist prior to this effort. Thus, the project was worthy of undertaking.

In building a rationale for the importance of the project, the authors gave considerable attention to position statements related to scientific literacy. Their argument was primarily that these various position statements offered over a few short years had in general grown more specific and ultimately offered a basis for developing an instrument the purpose of which would be to assess students' understandings of one aspect of scientific literacy, namely, the nature of scientific knowledge. Although this was an important argument from the perspective of literature of science education, it was done to the exclusion of reviewing literature related to the development of the "Rubba Model of Scientific Knowledge." Beyond stating that "A review was conducted of literature on the nature and philosophy of science..." the authors should have given some indication of the breadth and depth of this review.

Another review of the literature which might have been mentioned, if it were done, would be literature related to measurement instruments which did exist at the time this project was undertaken. That is, how does this newly developed instrument compare with, or complement, existing instruments in terms of elements within the various instruments, their purposes of existence, and the means of development? For example, is this instrument similar or dissimilar to instruments concerning attitudes about science? (Allen, 1959; Klopfer, 1966; Lowery, 1966) Does it address social issues in any way similar to that of other instruments? (Korth, 1968; Steiner, 1971) How does it compare to instruments which were developed to assess students' understandings of the nature of science? (Klopfer, 1963; Kimball, 1967, 1968) Is there any relationship of this instrument on understanding the nature of scientific knowledge and those related to understanding processes of science? (Welch, 1967; Tannenbaum, 1971; Wood, 1972)

This reviewer appreciated the step-by-step description of the development of the instrument. This is a difficult and time-consuming process which needs guidance. Surely, the authors have given other would-be instrument developers a solid prescriptive methodology which they might also use. Too often instruments are developed on something less than a theoretical basis. Such was not the case here. The Rubba Model of Scientific Knowledge served well the processes of writing items, refinement of draft instruments, and judgments made by the various panels of experts.

Questions might be raised with the approach used to establish reading level, especially in light of the variation which was found to exist in the reliability estimates for the instrument when used with general science students ($r_{kk} = 0.65$), biology students ($_{kk} = 0.74$), chemistry students ($r_{kk} = 0.74$), physics students ($r_{kk} = 0.77$), advanced chemistry students ($r_{kk} = 0.89$), college freshmen ($r_{kk} = 0.80$) and college philosophy of science students ($r_{kk} =0.88$). With regard to the reliability coefficients, it was not clear from the report if they represented split-half, odd-even, or any one of the other measures of internal consistency (Cronbach and Azuma, 1962).

The authors are to be commended for investigating the validity of the instrument--in particular as they stated it, the construct validity. However, it might be questioned if in fact the authors actually were not examining the concurrent or predictive validity instead of the construct validity. It will be recalled they used as a basis of their examination the hypothesis that philosophy of science students would score higher than would students with no formal history and philosophy of science backgrounds. In this situation a criterion has been adopted by which judgment will be rendered as to whether the instrument is valid, hence the examination is that of a criterion-related validity--concurrent or predictive. In this case since time is not an important issue, it would probably be best to describe it as concurrent validity.

The hypothesis posited for the test of validity clearly is directional and allows for a one-tailed test. Question is raised as to why in the summary table two-tail probabilities were reported. One additional note with regard to the statistical tests--the _t_ test was made between the two groups' test scores on each subscale and on the total score. The basis of doing this required the assumption that independence existed between the seven dependent measures. No evidence was offered that the subscales were indeed independent of each other; furthermore, one surely would not want to assume that independence existed between any one of the subscale scores and the overall test score. Hence, instead of using a series of univariate tests, possibly a multivariate test should have been used.

REFERENCES

Allen, H. "Attitudes of Certain High School Seniors Toward Science and Scientific Careers." _Science Manpower Project Monographs_. New York: Columbia University, 1959.

Cronbach, L. J. and H. Azuma. "Internal-consistency Reliability Formulas Applied to Randomly Sampled Single-factor Tests: An Empirical Comparison." _Educational and Psychological Measurement_, 22: 645-664, 1962.

Kimball, M. E. "Understanding the Nature of Science: A Comparison of Scientists and Science Teachers." _Journal of Research in Science Teaching_, 5(2): 110-120, 1967-1968.

Klopfer, L. S. and W. W. Cooley. "The Evaluation of Specific Educational Innovations." _Journal of Research in Science Teaching_, 1(1): 73-80, 1963.

Klopfer, L. S. "A Semantic Differential Approach to the Assessment of Students' Images of Science, Scientists, and Science Instruction." Paper presented at the annual meeting of the American Educational Research Association, Chicago, February, 1966.

Korth, W. W. "The Use of the History of Science to Promote Student Understanding of the Social Aspects of Science." Unpublished Ph.D. dissertation, Stanford University, Palo Alto, Calif., 1968.

Lowery, L. F. "Development of an Attitude Measuring Instrument for Science Education." _School Science and Mathematics_, 66(5): 494-502, 1966.

Showalter, V. M. "What Is Unified Science Education? (Part 5) Program Objectives and Scientific Literacy." Prism II(2): 3-14, 1974.

Steiner, R. L. "A Factor Analysis Study of the Attitudes of Oregon High School Seniors Toward Socially Significant Science--Related Issues." Unpublished Ph.D. dissertation, Oregon State University, Corvallis, Oregon, 1971.

Tannenbaum, R. S. "The Development of the Test of Science Processes." Journal of Research in Science Teaching, 8(2): 123-136, 1971.

Welch, W. W. and M. O. Pella. "The Development of an Instrument for Inventorying Knowledge of the Processes of Science." Journal of Research in Science Teaching, 8(1): 64-68, 1967.

Wood, R. L. "University Education Students' Understanding of the Nature and Processes of Science." School Science and Mathematics, 72(1): 73-74, 1972.

IN RESPONSE TO THE ANALYSIS OF

Rubba, P. A. and H. O. Andersen, "Development of an Instrument to Assess Secondary School Students' Understanding of the Nature of Scientific Knowledge" by Lawrence R. Gabel. Investigations in Science Education, 8 (4): 53-58, 1982.

Peter A. Rubba
Southern Illinois University

Hans O. Andersen
Indiana University

A recent issue of ISE contained Gabel's (1980) abstract and analysis of the "Development of an Instrument of Assess Secondary School Students' Understanding of the Nature of Scientific Knowledge" (Rubba & Andersen, 1977). As authors of the article, we found the abstract to be as accurate a summary of the process used to develop the "Nature of Scientific Knowledge Scale" (NSKS) as could be provided under the space limitation, and are appreciative of the compliments paid by the reviewer concerning the worth of the project and the systematic process used to develop the NSKS. We concede that greater information concerning the literature review which founded the Rubba Model of Scientific Knowledge could have been provided, and admit to the inappropriate statement of two-tailed t-test probabilities in testing the construct validity related hypothesis (though, the two t-values designated as not significant would remain not significant given the one-tailed probabilities).

However, the authors do not agree with the reviewer on a number of other points of critique. Several of these are a matter of judgment as to what should and should not have been included in the report. Others bring to question the reviewer's understanding of the process used to develop the NSKS and instrument development procedures in general.

Concerning the reviewer's desire for comparative information in the report on the NSKS and other instruments which measure aspects of understanding the nature of science, space limitations did not allow inclusion of information from the pre-NSKS-development review of existing instruments. A post-development discussion which compared the NSKS with instruments such as the "Nature of Science Scale" (NOSS) (Kimball, 1968), and the "Test on Understanding Science (TOUS) (Cooley & Klopfer, 1963), was not undertaken because the authors felt this type of comparison would best be made on an empirical basis. A review of the model statements founding the NSKS, NOSS, and TOUS does not make clear content similarities and differences. Concurrent validity studies need to be completed on the NSKS and other instruments which measure aspects of understanding the nature of science.

Some variation in NSKS reliability was anticipated over the large range of grade levels in which the instrument was tested (Nunally, 1970). Though the coefficient alpha reliability values on the instrument appear to be associated with respondent grade level, we doubt this is due to student reading level variance. As was stated in the article, "Because reading level formulas generally require a continuous sample of at least 100 words (Likert-type item statements do not fulfill this criterion), they could not be used...." The procedures employed were developed after consultation with two reading educators. It was their belief, and still is, that submission of the item statements to "...nine sixth grade students, of comparable reading ability as measured by the Iowa Test of Basic Skills..." was a more content valid method for determining item readability than is application of a reading formula. Given the upper level secondary students are very likely to be exposed in science class to issues in the philosophy of science, e.g., hypothesis testing, we believe item "interpreted" ambiguity to be the source of error responsible for the reliability coefficient variations.

The nature of the reliability coefficients reported in the article was questioned by the reviewer. Again, as was stated there, "coefficient alphas, r_{kk}," were reported. These reliability coefficients provide an assessment of internal consistency for an instrument composed of multi-point items. The Kuder-Richardson Formula 20 is a version of coefficient alpha for an instrument composed of dichotomous items (Nunally, 1967, pp. 196-197; pp. 550-551). Nunnally states that, "it (coefficient alpha) is so pregnant with meaning that it should routinely be applied to all new tests." (1967, p. 196).

In answer to the reviewer's question, "if in fact the authors were not examining the concurrent or predictive validity (of the NSKS) instead of the construct validity," our response is to pose a question for the reviewer. Concurrent to what; what was the external variable(s) considered to be a direct measure of understanding the nature of scientific knowledge with which NSKS administration results were compared? Criterion-related validity (concurrent or predictive) needs to be demonstrated for an instrument which is meant to provide a measure of a characteristic or behavior. In one sense all instruments are predictive; "they 'predict' a certain kind of outcome, some (past,) present or future state of affairs" (Kerlinger, 1973, p. 460). Nonetheless, understanding the nature of scientific knowledge is an abstract concept (defined by way of the Rubba Model of Scientific Knowledge). The overriding validity question associated with the NSKS was whether or not it measured the construct, understanding the nature of scientific knowledge.

Cronbach and Meehl (1966) describe five empirical methods of gathering evidence of an instrument's construct validity. The procedure referred to as "group differences" can be applied when an understanding of the construct allows one to anticipate that two groups will differ on a construct. The construct validity of the instrument can be tested directly by using the instrument to assess each group and then comparing the groups' scores. The instrument's ability to differentiate between the two groups can be evidence of its construct validity (p. 75). This procedure was the one used by the authors in employing the ex post facto design.

Finally, with regard to the suggestion that, "possibly a multivariate test should have been used" in the construct validity study, we are not able to identify what that technique might be. To our knowledge neither multiple regression, canonical correlation, discriminant analysis, nor factor analysis techniques could have been applied to the design used in order to elucidate NSKS construct validity.

REFERENCES

Cooley, W. W. & Klopfer, L.E. "The Evaluation of Specific Educational Innovations." Journal of Research in Science Teaching, 1(1): 73-80, 1963.

Cronbach, L. J. Essentials of Psychological Testing (3rd edition). New York: Harper & Row, 1970.

Cronbach, L. J. & P.E. Meehl, "Construct Validity in Psychological Tests." In C. Chase and H. Ludlow (Ed.), Readings in Educational and Psychological Measurement. Boston: Houghton Mifflin, 1966.

Kerlinger, F. N. Foundations of Behavioral Research (2nd edition). New York: Holt, Rinehart and Winston, 1973.

Kimball, M. E. "Understanding the Nature of Science: A Comparison of Scientists and Science Teachers." Journal of Research in Science Teaching, 5(2): 110-120, 1968.

Nunnally, J. D. Psychometric Theory. New York: McGraw-Hill, 1967.

Nunnally, J. C. Introduction to Psychological Measurement. McGraw Hill, 1970.

Rubba, [illegible]. A. & H.O. Andersen, "Development of an Instrument to Assess Secondary School Students' Understanding of the Nature of Scientific Knowledge." Science Education, 62(4): 449-458, 1977.

Wollman, Warren. "Controlling Variables: A Neo-Piagetian Developmental Sequence." Science Education, 61(3): 385-391, 1977.
Descriptors--*Abstract Reasoning; *Cognitive Development; Concept Formation; *Developmental Psychology; Educational Research; Elementary Secondary Education; *Learning Theories; *Logical Thinking; Science Education; Sequential Learning

Expanded abstract and analysis prepared especially for I.S.E. by A. W. Strickland, Indiana University.

Purpose

The purpose of this study was twofold: 1) to ascertain and magnify the difficulty in assessing the transition of learners from the concrete operational stage to the formal operational stage, and 2) to illustrate the impracticality of treating such data as dichotomous when an ordinal assessment provides a much more realistic view.

Rationale

In order to fully grasp the nature of this study it is necessary to examine the preceding study, "Controlling Variables: Assessing Levels of Understanding" (Wollman, 1977). From this earlier research the reader can gain insight into the establishment of the scale used in the present study and more clearly understand the nature of the inferences the students are attempting.

The author attempts to establish an instrument which may be used to distinguish between relatively concrete and relatively formal levels of logical dev^lopment. The author believes that the concept of controlling variables provides a nearly context-free area for the examination of the transition between concrete and formal operati ıal thinking.

Research Design and Procedures

Subjects. The Ss were from the same urban and suburban communities in the San Francisco area. The author indicated that the N's were the same as an earlier study (Wollman, 1977).

The author used the phenomenon of a sphere rolling down a grooved incline and striking a target sphere at the bottom, thereby sending it up another incline as the basis for his five-item test. The student responses were then evaluated according to a procedure described in Wollman's study (1977). The instrument was designed to assess students aged 11-18 years. The five-item test was administered to all students at the various grade levels. The scoring of the test also resembled the procedure used in Wollman's 1977 research.

Findings

The results of this research established a scale of difficulty for the five-item test, the difficulty from greatest to least followed the question sequence 1, 2, 5, 3, 4. The data implied a strong relationship between Question 5 and 3 (P = 0.94). The data indicate a gradual increase in percentage of correct responses as the grade le\ of the Ss increases. Moreover, the pass rates for all Ss on Question 4 was 80 percent; on Question 3, 62 percent; Question 5, 48 percent. The data for Questions 1 and 2 were not presented except as part of a contingency table comparing them with Question 5.

Interpretations

The author suggests that results such as these "are usually interpreted by developmental psychologists as implying the development of a single underlying ability." The author infers from the statement that this instrument may provide a continuum measure for concept development. He implies that Questions 3 through 5 should be less difficult because they ask for evaluations after the experiment has been tested, where Questions 1 and 2 require the synthesis of an experimental design.

The author indicates that the five-item test is only meant to be suggestive, serving only as a model for further research. The author also differs as to the question of instrument validity by referring to similar studies with varying content previously reported. In summary, Wollman makes the following analysis of the study:

> Since the concept of controlling variables is at the heart of the notion of valid empirical inference, and since empirical inference, the meaning of evidence, is both the test and the springboard of theory and conjecture in the social sciences as well as the natural sciences, the question sequence discussed here may suggest a generally useful technique for designing sequential learning experiences and assessment instruments consonant with the course of intellectual development as seen from a Piagetian viewpoint.

ABSTRACTOR'S ANALYSIS

The author's efforts in attempting to demonstrate a continuum of cognitive development between the concrete and formal operational stages is commendable. The rigor of his research must be impressive to fellow Piagetian researchers. It is also obvious that the author is pursuing an area of research which appears difficult to discuss within the limitations of a single article, and contains many aspects which may add to the confusion regarding Piagetian research rather than clarify.

One area of confusion may be unfair to the author as I am sure he intended for this article and the previous one (Wollman, 1977) to read in sequence. However, in examining this article the reader is constantly forced to return to the earlier one for clarification and understanding. The following may serve to illustrate the point:

1. The author never clearly identifies his sample size or distribution by grade level; you must, therefore, refer to Wollman (1977) for insight.

2. The author states, "To see how this general hypothesis might work, a three-item test was constructed to differentiate three levels of critical awareness" (page 386). Yet the test displayed in the article and used for data analysis is a five-item test. I inferred (perhaps incorrectly) that he created three additional items and added them to the two used in the previous study (Wollman, 1977).

3. Since the data for Questions 1 and 2 have been discussed in an earlier article (Wollman, 1977), the author avoids any meaningful discussion. This leaves the reader to search for any evidence that Question 1 and 2 were part of this study.

In the section described as "Method" the author makes the following statement: "Most but not all Ss were given these questions." One assumes he is referring to the five-item test. If so, why weren't all the students given the questions? If not, what does he mean? Additionally, the author indicates that the tests were given in groups and that each group, regardless of age level, took about ten minutes to complete the test. Does that mean that fourth graders and twelfth graders took the same amount of time to complete it? Why wasn't time considered a significant variable?

With regard to the instrument, it seems that when the researcher looks for certain written responses he relies very heavily on their verbal ability. Many of the fourth and fifth graders I showed the questions to could not read them. And even after I read the questions to them, few understood what was being asked.

The author interprets the data to illustrate a continuous learning scale. Perhaps that inference should be reserved until the validity and reliability of the instrument has been clearly established. The author does not cite validity or reliability data for this instrument.

REFERENCE

Wollman, Warren. "Controlling Variables: Assessing Levels of Understanding." Science Education, 61(3): 371-383, 1977.

IN RESPONSE TO THE ANALYSIS OF

Wollman, Warren. "Controlling Variables: A Neo-Piagetian Developmental Sequence" by A. G. Strickland. Investigations in Science Education, 8 (4): 62-66, 1982.

by

Warren Wollman
Arizona State University

Professor Strickland has written a thoughtful commentary, but some points of confusion exist, an unavoidable circumstance due to the brevity of research papers. I want to try to clear up the most important points I tried to make in the paper and in its immediate predecessor in the same journal (1977a,b).

Rather than "magnify the difficulty in assessing the transition" from concrete to formal, I wished only to show that the said transition is in point of fact more complex than Piaget suggests, more difficult to characterize than most science eudcation researchers appear to realize or accept. The data speak for themselves. (Parenthetically, Professor Inhelder agrees wholeheartedly with me that her research on formal reasoning, now almost 30 years old, should be viewed as only a first attempt to shed light on a very difficult matter to study, namely the nature of adolescent reasoning. She unhesitatingly agrees that new criteria must be found for describing the transition from concrete reasoning to more mature forms--personal communication.)

My two papers illustrated a general way for assessing levels performance on tasks associated with the formal stage. Levels differed according to abilities to meet demands on amount, type, and organization of information. My own feeling at that time was that Piaget and those who take him at face value, more or less, grossly oversimplify descriptions of task demands and thus fail to observe that unacceptable performance can be due to many and diverse factors. By gaining a clearer understanding of task demands, we obviously can gain a clearer understanding of what needs to be done, by teacher and student, to meet task demands. If one simply accounts for poor performance as being due to lack of intellectual maturation, then one tacitly admits to being unable to substantially improve performance over a time scale which is short of a maturational scale.

Strickland feels that I tried to provide a "context-free areas for the examination of the" concrete-formal transition. If anything, I tried to show how foolish it is to ignore context and that formal reasoning is not context-free. I simply tried to design tasks whose context would not be confusing for reasons uninteresting to educators, e.g., confusing because the task was poorly worded or confusing because the materials were alien to the students' experience. At most, I aimed for the development of a context-free method for assessing levels of performance. Such a method at its best would, in my judgment, provide a way to specify and quantify the amount of information required, as well as quantify the attentional demands of organizing and operating upon that information.

I was unaware at the time of the research of Professor Juan Pascual-Leone, who had made a breakthrough in this direction (1969 and, more recently, 1980). His students have since gone on to apply his theoretical ideas along the lines I only vaguely perceived at that time (see Case 1978a,b, for reviews).

Strickland does well to raise an issue as to whether or to what extent verbal ability should be used to measure intellectual level. There is no doubt that the two are related, but it is not at all clear how they are related. By using a written test format, I probably overemphasized certain language skills, though I doubt whether anyone can tell how much. Assuming that verbal ability was overemphasized, my data would give orderly conservative age norms, but the sequence of performance levels would remain much the same. In other words, the sequence is a reliable aspect of my data.

As to other notions of reliability and validity, my pilot studies convinced me that the data were reliable at least for average students (very bright students might well do better a second time because of their propensity for reflecting upon and making sense of intellectual challenges). Written tests followed by interviews gave very consistent results. The question of validity is quite another matter. I do not believe that Piaget and Inhelder's experiments are completely valid measures of adolescent reasoning. If I did, I would not have gone out of my way to design new measures. Moreover I now know that Professor Inhelder is open to reconsidering how to empirically delineate the development of adolescent reasoning. It would be narrowminded indeed to estimate the validity primarily on the basis of similarity with Piaget's highly selective reading of Inhelder's impressive store of protocols. Piaget informed me that he sampled data in order to illustrate his theoretical ideas, that his characterizations of performance were not descriptions of average or typical behavior, and that his stage descriptions were essentially definitions or hypotheses awaiting confirmation. I cannot imagine why he did not make this clear when he wrote The Growth of Logical Thinking. I discuss the validity of Piaget's work at great length elsewhere (1978).

Finally, Strickland raises questions of method concerning the length of the tests, the time allotted for completion, and the brevity of the reviewed paper. Interested parties are respectfully urged to (a) write to me for clarifications, and (b) read the paper immediately preceding the reviewed paper in the same journal. Better still, interested parties should peruse the psychology journals such as Child Development, Cognitive Psychology, and Developmental Psychology if they wish to update their thinking on the psychological aspects of what Piaget calls formal reasoning. In particular, see the work of Case, cited above, and Siegler (1976, 1980).

REFERENCES

Case, R. "Intellectual Development from Birth to Adulthood: A Neo-Piagetian Reinterpretation." In R. S. Siegler, Editor, Children's Thinking: What Develops? Hillsdale, NJ: Lawrence Erlbaum Associates, 1978a.

Case, R. "Piaget and Beyond: Toward a Developmentally Based Theory and Technology of Instruction." In R. Glaser, Editor, Advances in Instructional Psychology. Hillsdale, NJ: Lawrence Erlbaum Associates, 1978b.

Pascual-Leone, J. Cognitive Development and Cognitive Style. Unpublished doctoral dissertation, University of Geneva, 1969.

Pascual-Leone, J.; D. Goodman; P. Ammon; and I. Subelman. "Piagetian Theory and Neo-Piagetian Analysis as Psychological Guides in Education." To appear in J. M. Gallagher and J. Easly (Eds.), Knowledge and Development (vol. 2): Piaget and Education. New York: Plenum Publishing Co., 1980.

Siegler, R. S. "Three Aspects of Cognitive Development." Cognitive Psychology 8: 481-520, 1976.

Siegler, R. S. Developmental Sequences Within and Between Concepts. To appear as a Society for Reserach in Child Development monograph, 1980.

Wollman, W. "Controlling Variables: Assessing Levels of Understanding." Science Education 61(3): 371-383, 1977a.

Wollman, W. "Controlling Variables: A Neo-Piagetian Developmental Sequence." Science Education 61(3) 385-391, 1977b.

Wollman, W. Developmental Implications of Science Teaching: Early Adolescence. Columbus, OH: ERIC Clearinghouse for Science, Mathematics and Environmental Education, The Ohio State University, 1978.

Wheeler, Alan E. and Heidi Kass. "Student Misconceptions in Chemical Equilibrium." Science Education, 62(2): 223-232, April-June 1978.

Descriptors--*Chemical Equilibrium; *Chemistry; *Cognitive Development; *Educational Research; Learning Difficulties; Learning Theories; Science Education; *Science Instruction; Secondary Education; *Secondary School Science

Expanded abstract and analysis prepared especially for I.S.E. by Richard M. Schlenker, Maine Maritime Academy.

Purpose

The investigators' major purpose was to determine what relationships, if any, existed between Ss reasoning abilities, their achievement in high school chemistry and the misconceptions they possess concerning chemical equilibria. Although they did not formulate specific hypotheses concerning the outcomes of their research, the investigators did pose several research questions.

1. What is the nature of Ss misconceptions about chemical equilibrium?
2. What is the extent of Ss misconceptions about chemical equilibrium?
3. What is the degree to which six major misconceptions are related to chemistry achievement? The misconceptions investigated were:
   a) Mass vs. concentration--the inability to distinguish between the concepts of mass and concentration.
   b) Rate vs. extent--the inability to distinguish between the rate which a reaction proceeds and how far that reaction will proceed.
   c) Constancy of the equilibrium constant--the uncertainty about when the equilibrium constant was a constant.
   d) Misuse of Le Chatelier's principle--the application of the principle--type of reasoning in inappropriate situations.
   e) Constant concentration--the inability to conceptualize that certain substances display a fixed or constant concentration in certain chemical reactions.
   f) Competing equilibria--the inability to consider all possible factors affecting the equilibrium condition of a chemical system.

4. What is the degree to which the misconceptions are related to performance on two tasks each involving the mixing of five colorless solutions in various combinations to produce a colored solution?

Rationale

The research and procedures were forged within the Piagetian concrete-formal conceptual framework and were based upon the following assumptions:

1. That the use of n-by-n combinations of colorless solutions in a systematic way to produce a color and the understanding that said systematic use are demonstrations of an ability to apply formal operational reasoning structures.
2. That the formal mode of reasoning concerning the existence of color in solutions, which bases the establishment of color upon a combination of factors, leads individuals to conceptualize that the establishment of color is brought about by the reactions between solutions.
3. That concrete operational thinkers search for the reasons why color appears, following the mixing of solutions together, in one or another of the solutions that were mixed together without attributing said cause to the union of solutions.
4. That individuals who do not use formal operational reasoning structures differ in degree as to their ability to attribute the proper cause to the appearance of color following the mixture of solutions.
5. That the closer individuals are to being formal thinkers the quicker they are to attribute the appearance of color to the mixture of solutions.
6. That the closer individuals are to being formal thinkers the more systematic they are when mixing a series of solutions together to produce a color.

The study was based upon Inhelder's and Piaget's (1958) description (identity, negation, reciprocity, correlativity; hereafter referred to as INRC) of the way adolescents manipulate data derived from experiments as described by Flavell (1963) and clarified by Parsons (1960).

Research Design and Procedures

Sample: The sample consisted of 99 twelfth-grade chemistry students from four chemistry classes. Sixty-four percent of the students were males and 36 percent were females.

Instruments: Each of the following instruments were administered to all Ss involved in the study:

1. The Misconception Identification Test (MIT). This instrument required subjects to predict the effect of changing variables upon the equilibrium conditions of selected chemical systems. It was designed to investigate the misconceptions listed as 3A-F in the PURPOSE section above.
2. Chemistry Achievement Test based upon Chapters 7-10 of the CHEM Study text.
3. Two tasks, each involving the mixing of five solutions in various combinations to produce a color.
4. A written test involving INRC transformations.

Administration: The instruments were administered following the completion of class work upon relev nt CHEM Study chapters. All of the instruments were administered over a period of approximately one week.

Data Manipulations: Several mathematical techniques were used to evaluate the data. Chi-square analyses were used to evaluate the relationship between the performance portion and the misconception portion of the MIT, the degree of independence between numbers of misconceptions and cognitive level and the independence between achievement and numbers of misconceptions. Stepwise multiple regression analysis was used to predict MIT scores based upon chemistry achievement, solution combinatorial task and INRC scores and to predict Chemistry Achievement Test scores from solution combinatorial task and INRC scores. Intercorrelations between all scores on all instruments were also computed.

Findings

The following items represent the major findings:

1. Eighty-two percent of the Ss possessed three or more misconceptions.
2. In terms of cognitive functioning, 3 Ss were early concrete, 24 Ss were late concrete, 61 Ss were early formal, 11 Ss were late formal.
3. Scores on the two sections of the MIT were related at $p < 0.01$.
4. MIT scores were significantly related to INRC scores.
5. A large portion of the observed variance in the MIT (performance section) was attributable to chemical solution task score variations.
6. A large portion of the observed variation in the MIT (misconception section) was attributable to variation in Chemistry Achievement Test scores.
7. The relationship between number of misconceptions and cognitive level was significant at $p < 0.01$.
8. Mass vs. concentration and rate vs. extent were related to cognitive level at $p < 0.05$.
9. 58.7 percent of Chemistry Achievement Test scores were predictable using a combination of INRC and chemistry solutions test scores.
10. Consistency of the equilibrium, misuse of Le Chatelier's principle and competing equilibria were related to chemistry achievement at $p < 0.05$.

Interpretations

The authors concluded:

1. That inability to control variables in chemical equilibrium problems probably affects demonstrated achievement.
2. That, prior to introducing students to chemical equilibrium, their cognitive levels should be assessed.
3. That instruments used in this study to assess cognitive level are adequate in the area of chemistry.

The following suggestions were made, based upon the outcomes of the study:

1. Concrete students should benefit from laboratory sessions involving chemical equilibrium.
2. The use of programmed materials involving chemical equilibrium should aid both concrete and formal students in their understanding of equilibrium concepts.
3. A large number of qualitative and quantitative examples of chemical equilibria be made available to students.
4. The use of graphical representations of chemical equilibria will aid students in understanding the time--concentration concept.

ABSTRACTOR'S ANALYSIS

Chemistry by its very nature is extremely abstract; therefore, it is often assumed students must possess high IQ's or, so to speak, be bright (have a high level of ability) if they are to demonstrate high levels of competency in chemistry study. That the former portion of this statement is true and the second portion is patently obvious and defensible in the perceptions of many chemistry instructors, who lack an understanding of learning theory, is made a matter of fact by those whose teaching methods require students to have high levels of abstract or writing about its outcomes, one is left to wonder whether outside influences upon the subjects in this study might have confounded the results of the investigation. For example, what is the relationship between observed outcomes and students' backgrounds and participation in other courses at the time of the study? It seems quite likely that subjects enrolled in mathematics courses at the same time might have tested more formally than others (assuming transition from the use of concrete to formal reasoning structures can at least be encouraged on a temporary basis). This as well as previous course background, family background and so on might be controlled at least partially, using random sampling techniques.

Generalizability of the results beyond the sample is at best difficult because the information provided concerning the sample is weak. I have alluded to this in the previous paragraph. However, here I must ask from whence the sample came: the U.S., Canada (at the time of writing, prior to July 1977, the authors were at the University of Alberta), small community, large affluent community, bilingual community and so on? In addition, one might ask when the study was conducted.

The authors' suggestion that concrete instruction will aid concrete students is supported by Sheehan (1970). However, the suggestion that plotting changes in concentration over time may help students to "concretely visualize" what is thought to happen requires clarification. At this stage in our understanding "concretely visualize" is a contradiction. In fact, the ability to visualize appears to be a trait inseparable from abstract reasoning ability (Arnheim, 1969; Schlenker, 1977; Wallach, 1961). The suggestion is made and supported in the literature that concrete reasoners lack the ability to visualize to a high degree.

Finally, the authors suggest by the use of early concrete, late concrete, early formal and late formal that actual stages of reasoning exist whereas Piaget (1972) himself suggests a reasoning continuum likened to the stages of embryogenesis.

Withstanding the criticisms of the study made herein, the authors are to be applauded for providing another valuable link in our understanding of learning in the physical sciences.

Suggestions for Further Research

Perhaps the one well-supported suggestion for further research coming out of this paper is that student functioning in both chemistry and physics courses should be evaluated amongst students taking the courses simultaneously. The objective should be to ascertain whether there is a differential between the physics achievement-cognitive ability relationship and the chemistry-cognitive ability relationship. It might be hypothesized that students having taken high school physics prior to chemistry or the converse might have an advantage over those students taking their first physical science course. Such an inference might be made if it appeared a differential did not exist. This inference of course would also require the evaluation of students having taken physics before chemistry and so on, and might further suggest students entering one course without having the other as a background to be at a disadvantage.

Withstanding the criticisms of the study made herein, the authors are to be applauded for providing another valuable link in our understanding of learning in the physical sciences.

Suggestions for Further Research

Perhaps the one well-supported suggestion for further research coming out of this paper is that student functioning in both chemistry and physics courses should be evaluated amongst students taking the courses simultaneously. The objective should be to ascertain whether there is a differential between the physics achievement-cognitive ability relationship and the chemistry-cognitive ability relationship. It might be hypothesized that students having taken high school physics prior to chemistry or the converse might have an advantage over those students taking their first physical science course. Such an inference might be made if it appeared a differential did not exist. This inference of course would also require the evaluation of students having taken physics before chemistry and so on, and might further suggest students entering one course without having the other as a background to be at a disadvantage.

REFERENCES

Arnheim, R. Visual Thinking. Berkeley: University of California Press, 1969.

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

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

Parsons, C., B. Inhelder, and J. Piaget. The Growth of Logical Thinking: A Logician's Viewpoint." British Journal of Psychology, 51: 75-84, 1960.

Piaget, J. "Intellectual Evolution from Adolescence to Adulthood." Human Development, 15: 1-12, 1972.

Schlenker, R. M. "The Relationship Between Lowenfeld's Visual-Haptic Perceptual Continuum and Piaget's Concrete-Formal Intellectual Aptitude Continuum in College Students." Unpublished doctoral dissertation, University of Maine, 1977.

Sheehan, D. J. "Effectiveness of Concrete and Formal Instructional Procedures with Concrete- and Formal-Operational Students." Doctoral dissertation, University of the State of New York at Albany, 1970. Dissertation Abstracts International, 1970, 31: 2748A.

Wallach, H. "Some Considerations Concerning the Relation Between Perception and Cognition." In Henle, M. (Ed.), Documents of Gestalt Psychology. Berkeley: University of California Press, 1961.

IN RESPONSE TO THE ANALYSIS OF

Wheeler, A. E. and H. Kass. "Student Misconceptions in Chemical Equilibrium," by Richard M. Schlenker. Investigations in Science Education, 8 (4): 70-78, 1982.

by

A. E. Wheeler
Brock University

Professor Schlenker's reflective comments and analysis on our earlier study, "Student Misconceptions in Chemical Equilibrium" in I.S.E. center on two areas: one concerning a felt need for further elaboration on the nature of the sample involved in the study and further background details surrounding the Ss and the second with capabilities of the so-called "concrete thinker" in chemistry and the possible influence of certain instructional strategies designed to enhance the capabiliteis of such students.

To answer Professor Schlenker's questions in the first area, the study was conducted on 99 Grade Twelve CHEM Study students in the spring semester of 1973 following completion of instruction of the relevant portions of the course dealt within the Chemistry Achievements Test (CHAT). The four classes involved were drawn from two high schools in a large urban Canadian center. It is of interest to note that all Ss in the investigation were also enrolled in a common mathematics course at the time. Grade nine Co-operative School and College Ability Test (SCAT) scores (Form 3A) were also available for all subjects. Professor Schlenker's concern that varying student backgrounds in mathematics instruction may have confounded the results of the investigation was, in this sense, partially controlled for. The fact that neither the verbal or quanititative SCAT scores entered the regression equation would tend to support this contention. However, Professor Schlenker's suggestion that mere exposure to mathematics instruction facilitates cognitive growth, even on a temporary basis, is one which is open to debate. Findings in a later study on proporational reasoning in chemistry (Wheeler, 1976) support this apparent lack of transfer between the application of mathematics in a traditional context and application in a chemistry context.

In his analysis Professor Schlenker suggested that the use of the phrase "concretely visualize" which was offered in connection with certain graphical representations which may serve as possible vehicles to enhance formal thought was in itself a contradiction. While a portion of this contention may be semantic, the essence would appear to reside in the vary nature of concrete and formal thought as delineated by Piaget. According to Professor Schlenker the ability to visualize appears to be a trait <u>inseparable</u> from abstract reasoning ability. This, we would suggest, is in contradiction to the reasoning continuum referred to in Professor Schlenker's comments which we fully support.

While it may well be true that the concrete reasoner lacks the ability to visualize to a high degree, he surely can and must be able to visualize to some degree. Our suggested instructional device was offered only in order to tap and develop this ability.

REFERENCE

Wheeler, Alan E. The Role of the Proportionality Schema in Introductory High School Chemistry. Unpublished Doctoral dissertation, The University of Alberta, 1976.