DOCUMENT RESUME

ED 324 200 SE 051 579

AUTHOR	Blosser, Patricia E., Ed.; Helgeson, Stanley L., Ed.
TITLE	Investigations in Science Education. Volume 14, Number 4, 1988.
INSTITUTION	Ohio State Univ., Columbus, Ohio. Information Reference Center for Science, Mathematics, and Environmental Education.
PUB DATE	88
NOTE	101p.; For previous edition, see ED 309 037.
AVAILABLE FROM	SMEAC Information Reference Center, 1200 Chambers Road, Room 310, Columbus, OH 43212 ($2.75 single copy).
PUB TYPE	Collected Works - Serials (022)
JOURNAL CIT	Investigations in Science Education; v14 n4 1988
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	Attitude Measures; Biology; Chemistry; Cognitive Development; *College Science; Educational Assessment; Elementary Education; *Elementary School Science; Energy; Higher Education; Misconceptions; Physics; Process Education; Questioning Techniques; Research Methodology; Science Curriculum; Science Education; *Science Instruction; Science Tests; Secondary Education; *Secondary School Science; Student Attitudes; Thinking Skills
IDENTIFIERS	*Science Education Research; *Wait Time

ABSTRACT

This volume includes abstracts and abstractors' critiques of 15 published research reports related to elementary, secondary and college science teaching and learning. Topics include: (1) assessing attitudes toward energy conservation; (2) persuasive communications; (3) questioning techniques; (4) teachers' verbal exposition on student participation in biology classes; (5) verbal behaviors; (6) cognitive skills; (7) reading micrographs; (8) writing assignments in general chemistry; (9) prelaboratory procedures in chemistry; (10) short courses in science; (11) wait time; (12) science teaching learning attributes of students; (13) classroom evaluation materials; (14) teaching energy concepts; and (15) children's understandings of cause and effect relationships. The responses of three authors to the critiques of their research are reported. (CW)

INVESTIGATIONS IN SCIENCE EDUCATION

National Association for Research in Science Teaching

ERIC Clearinghouse for Science, Mathematics, and Environmental Education

INVESTIGATIONS IN SCIENCE EDUCATION

Volume 14, Number 4, 1988

Published Quarterly by

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

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

Volume 14, Number 4, 1988

NOTES FROM THE EDITOR:

Issue 4 of ISE is another collection of miscellaneous studies, although there are some clusters within the issue. Two articles by Koballa focus on attitude research. These are followed by several articles that relate in one way or the other to achievement: Riley studied the relationship of a teacher's choice of cognitive level of questions and wait-time to student achievement; Okebukola and Ogunniyi examined different instructional strategies and student achievement; Ogunniyi looked at teachers' verbal behavior during instruction; and Hudson considered the effects of diagnostic feedback on student performance in physics. Also in the achievement cluster are critiques of articles by Johnson and Lockard on student ability to read micrographs, by Horton et al on using writing assignments in college chemistry, by Isom and Rowsey on the use of small-group instruction in chemistry, and by Stanley and Stanley on the effects of intensive instruction in science on highly able youth.

Teacher wait time and its effects on several outcome variables are described in an article by Tobin. Harty and Beall examined the affective domain and how it relates to student success in science. Atwood et al studied the development of criterion-referenced tests for assessing the achievement of elementary students working with SCIS materials. Urevbu investigated concept learning among Nigerian primary school children. Rudnitsky and Hunt examined strategies used by children to solve a complex problem while they were using a computer.

Also included in this issue are two responses to critiques. Ogunniyi has provided a response to Craven's critique of his investigation of classroom interaction, and the Stanleys have responded to Perry's critique of their report of the progress of talented students in fast-paced summer science courses.

This issue also marks, at least for the time being, the end of production of INVESTIGATIONS IN SCIENCE EDUCATION due to financial constraints. This has been a pleasant task, and we are most appreciative of the time and efforts of those persons who have served as reviewers as well as those authors who have availed themselves of the opportunity to respond to critiques.

Patricia E. Blosser
Editor

Stanley L. Helgeson
Associate Editor

Koballa, Thomas R. Jr. "Designing a Likert-Type Scale to Assess Attitude Toward Energy Conservation: A Nine-Step Process." Journal of Research in Science Teaching, 21 (7): 709-723, 1984.

Descriptors--*Attitude Measures; Elementary Education; *Energy Conservation; *Inservice Teacher Education; *Preservice Teacher Education; Science Education; *Teacher Attitudes; *Test Construction; Test Reliability; Test Validity

Expanded abstract and analysis prepared especially for I.S.E. by Ronald D. Simpson, University of Georgia.

Purpose

The purpose of this investigation was two-fold. First, the author sought to develop a valid and reliable Likert-type scale that would measure attitudes of teachers toward energy conservation. Second, it was the author's aim to discuss the nine step process that was used in developing the instrument. Toward both of these ends, the author drew from important literature in social psychology and education in documenting the rationale for the procedures used in this study.

Rationale

The author points out the importance of being able to measure attitudes. He also discusses major approaches to attitude measurement and ends up sketching four prominent direct measurement techniques that have been used and documented in the research literature. He goes on to validate the fact that the Likert-type scale is viewed today as possessing several advantages over the other direct, self-report techniques. He, therefore, demonstrates the efficiency of using this method for measuring the important topic of attitude toward energy conservation among teachers.

Equally important to the rationale of this study is the fact that researchers in science education need to know the appropriate steps to take when developing instruments to measure attitude toward a given psychological object. This study addresses an important area of research methodology by bringing together key information from the literature and demonstrating how it can be used in a practical situation.

Research Design and Procedure

The investigator covers nine steps for developing a Likert-type scale for measuring attitudes. The first step involved assembling a pool of items consisting of 34 moderately positive and 37 moderately negative statements about energy conservation. These statements were drawn from conversations with preservice elementary school teachers and assumed that both positive and negative attitudes existed toward the construct being measured.

The 71 statements comprising the original pool were screened by an expert panel of three. The judges used the screening criteria suggested by Edwards (1957), which has become standard in Likert scale development. For step 3, the remaining pool of 60 items were piloted with 93 preservice elementary teachers. The teachers were not only asked to respond using the typical five point agree-disagree format, but were asked to comment on items that they felt were unclear or ambiguous.

Step 4 included analyzing the responses from these subjects using the Likert Attitude Scale Analysis Main Program (LIKRT) as developed by Kohr (1973) at The Pennsylvania State University. A summary table was printed with the frequency and percentage of occurrence of each response choice for both the upper (most positive) 27% and lower (least positive) 27% of respondents. Item means and standard deviations were calculated for each group. Using this method the two contrast groups are compared with respect to mean scores, using a t-test. Ultimately the LIKRT procedure assesses the potential contribution of each item as judged by discrimination index, bipolar data, and neutral data.

In Step 5 the investigator selected 21 items that met the specific criteria of Step Four. The coefficient alpha for the 21 item scale was calculated at 0.84, an estimate suggesting that this instrument possessed reasonable internal consistency. In Step 6 the investigator assessed the attenuation of reliability by administering the 21 item scale to a total of 203 preservice elementary school teachers. No statements were found to be unclear or ambiguous and the coefficient alpha for this sample was 0.82, an attenuation of only .02. In Step 7 two statements that failed to meet neutral response rates and bimodal distribution were eliminated.

In Step 8 the coefficient alpha for the 19 item scale was calculated to be 0.81. In Step 9 validation procedures were followed wherein the 203 preservice teachers also responded to the 43 item Opinion Subtest of the Energy Inventory (BSCS, 1977). Point-biserial correlation coefficients equal to 0.59 and 0.57 were found between the Opinion Subtest and the revised 19 statement scale and the 21 statement scale, respectively. The investigator offers this convergent validation operation as evidence that the 19 and 21 statement scales do measure attitude toward energy conservation.

Interpretations

The author concludes that (1) a 19-statement scale targeted for use with preservice and inservice elementary school teachers was developed that measures attitude toward energy conservation and (2) the nine step model for the construction of a Likert-type attitude scale was demonstrated in useful manner. The investigator also reported that a point-biserial correlation coefficient of 0.89 was calculated for the scores of 85 preservice elementary school teachers over a three-week time lapse, further suggesting reliability over time.

The author is cautious in concluding that this 19-item instrument is a final product. He suggests that it should still be compared with other instruments when attempting to measure attitude toward energy conservation. He also suggests "monitoring and modifying" this instrument, particularly when it is used with populations other than preservice and inservice elementary teachers. The nine-step method for developing Likert-type attitude scales appears to be a useful mechanism for implementing these "monitoring" and "modifying" activities that should accompany the development and refinement of all affective measures.

ABSTRACTOR'S ANALYSIS

This study is unique in that it makes two parallel contributions; one, the production of a promising instrument for measuring attitudes toward energy conservation and, the other, advancing a process that should be quite useful to other researchers who in the future construct Likert-type instruments for attitude assessment. The first contribution is an end in itself. Knowing how elementary school teachers feel toward energy conservation and knowing how to measure changes in attitude is an important capability for the science education researcher, particularly one interested in the important topic of environmental education.

The 19-item instrument developed here appears a suitable one for measuring what the investigator claims and, as just stated, this is a valuable addition to our knowledge base in science education. The author is careful not to make undue claims regarding the instrument and is thoughtful to encourage others to join in the "monitoring" and "modification" process of this and other instruments.

The second contribution of this study is the most important one, in my opinion. This researcher has carefully and correctly chronicled the history of direct, self-report techniques used for attitude measurement, beginning with Thurstone's work and ending with the commonly agreed upon premise that the Likert-type scale offers the most parsimonious paper and pencil method for measuring group attitudes. The review of literature in this paper is authoritative and concise. Furthermore, important concepts are communicated in a clear and logical fashion.

Of course, the unveiling of the LIKRT model for developing a Likert-type scale is a specific contribution. The Likert Attitude Scale Analysis Main Program was developed at The Pennsylvania State University and appears to be a useful tool for assisting with the development of such scales. Though most of the components of the model are procedures known and accepted by attitude researchers, the integrated procedure described here results in a process that is both useful and novel. This should aid many instrument developers in the future and, thus, give science educators yet another medium through which to communicate and compare notes.

This investigation was well-written and, as stated before, was based on a well-conceived body of literature that is generally regarded as the theoretical base from which most of our practices in attitude research emanate. In short, the author offers us a blueprint to use in the future development of attitude instruments.

There are only minor points to make with regard to areas where improvement could have been made. In most cases the author acknowledged these shortcomings. One example is that an initially larger item pool would have possibly led to more appropriate items for the final version of the instrument. While mentioning item selection I might add that though I thought most of the items were quite good I did spot a few that I wondered about in terms of whether or not they actually measured "attitude" or "feeling." One example is the item "Conservation is presently a plausible alternative to the energy problem confronting us." I do not see this item eliciting much feeling from students nor from a face validity point of view do I see it discriminating to any extent between commitment and apathy. On another point, the item "Drivers who disobey the 55 MPH speed limit should be fined heavily" was listed as a negative item. I am sure this is an error since agreeing with this item would imply a positive attitude toward conservation. One additional point about this item is that "55 MPH" dates the item. The general term "speed limit" or "current speed limit" would have been a better way of writing this item for longevity purposes. Another item written in a negative posture was "If the price of gasoline increases, I will use less." One who is positive toward conservation would presumably disagree with that statement, yet I wonder if agreement with this statement can be interpreted as being negative toward conservation. The items in general are well-crafted and from examination appear to address important aspects of energy conservation.

The author mentions that perhaps a larger number of items initially would have been an advantage. I agree. Also, I would suspect that having a somewhat larger and more diverse panel of experts would have contributed positively to this process. A classroom teacher, a social psychologist and perhaps one or two other professionals in addition to the "three educators who had had previous experience with Likert scale construction" would have added breadth to this process. Also, the initial population of 93 subjects used to pilot test the 60 statement version of the instrument was small. Some psychometricians suggest that there be two to three times as many subjects as there are items when doing reliability estimates and other related analyses. There was some confusion, by the way, as to whether the initial instrument to be tested contained 71 or 60 trial statements. Some of the tables in this report, likewise, lacked a clear description. One final suggestion I would make is that Step 9 which dealt with validation procedures could have been expanded. This obviously is an extremely important aspect of attitude scale development and more discussion on this would have been helpful. Likewise, more information about the initial development of the LIKRT procedure would have strengthened the paper.

This report represents an important contribution to the literature in science education. This very capable investigator does a superb job in building a theoretical and historical base and then introducing a nine-step technique for developing a Likert-type scale for measuring attitudes towards energy conservation. The article is well written and authoritative. The investigator was both objective and open by encouraging others to join in the "monitoring" and "modification" process of using this and other instruments designed to measure attitudes. As we refine our methods for measuring attitude and other related constructs our potential for understanding more clearly how feelings relate to cognition will be raised. As this occurs, we will be able to communicate with a larger audience. This author, through this investigation, has demonstrated in a commendable fashion how this process can and should work.

Koballa, T. R., Jr. "Changing Attitudes Towards Energy Conservation: The Effect of Development Advancement of the Salience of One-Sided and Two-Sided Persuasive Communications." Journal of Research in Science Teaching, 21 (6): 659-668, 1984.

Descriptors--*Attitude Change; Cognitive Development; *Communication (Thought Transfer); Conservation Education; Elementary School Teachers; *Energy Conservation; Higher Education; *Persuasive Discourse; *Preservice Teacher Education; Science Education; *Teacher Attitudes

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

Purpose

The purpose of this work was to test the effectiveness of persuasive communications designed to fit the developmental character of the receiver's energy-relevant cognitive system.

Specifically, the study addressed the following two questions:

1. Are one-sided or two-sided systematically designed communications towards energy conservation more effective in persuading non-developmentally advanced and developmentally advanced preservice elementary teachers respectively?
2. Do positive attitude gains between pre- and post-tests, if any, dissipate within four weeks following the treatment of either one or the two of these groups?

Rationale

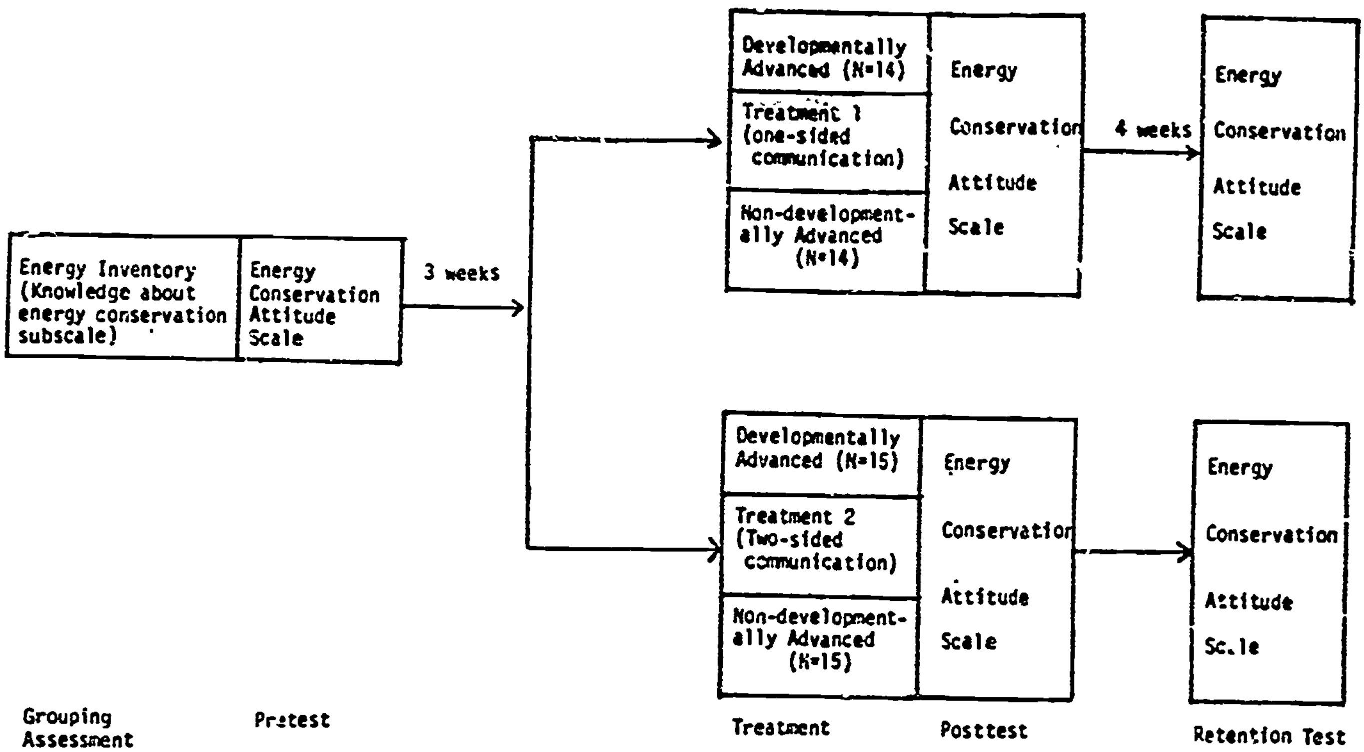
Although persuasive communication can be invoked as a model means for attitude change in science education (Shrigley, 1978), the extent to which the information/knowledge component of class instruction can (if at all) contribute to meaningful attitude modification -- leading eventually to actual behavior modification -- is a long standing controversial issue (Inkso, 1967). It appears to be agreed upon that such meaningful attitude modification to be achieved by a credible class teacher operating within a framework of consistent and systematic curricular design is contingent on the application of an appropriate multidimensional instructional model (Zoller & Maymon, 1989).

In accepting (a) the constructivism point of view concerning the personal (cognitive) construct which underlies human behavior; (b) Cronbach's challenge of designing the treatment to fit groups of students with aptitude patterns, not the average person; and (c) the persuasive communication model for attitude change; the theoretical undergirding for the study as well as the contextual framework within which the investigation was conducted, have been established. Thus, the underlying assumption is that the effectiveness of a one-sided persuasive communication containing only favorable arguments regarding the issue (energy conservation in this case), versus a two-sided persuasive communication containing both favorable and unfavorable arguments regarding the same issue, should depend upon the relevant information available to the receiver with respect to the attitude domain.

Research Design and Procedure

The 58 participating pre-service elementary teachers categorized as developmentally advanced and non-developmentally advanced--based on their scores on the "knowledge about energy conservation subscale" of the Energy Inventory (BSCS, 1977),--were randomly assigned to either the one-sided or two-sided communication. The former was a video-taped persuasive message undergrided by seven belief statements supportive of energy conservation. The latter was a message identical to the former in every respect except for the incorporation of counter-arguments within the body of the communication. Immediately following the exposure to the communications and four weeks later (pre- and post-tests) the modified Energy Conservation Attitude scale (Kobaila & Shrigley, 1983) was administered to all subjects. A schematic representation of the research design (as summarized by the researcher) is given on the next page.

Comparisons between the mean difference scores, posttest - pretest of developmentally and of non-developmentally advanced persons exposed to the one-sided and the two-sided communication respectively, as well as additional similar comparisons, were made in order to summarize and analyze the data.



Findings

The major findings of the study were as follows:

1. The two-sided communication was more effective in changing attitudes towards energy conservation immediately following treatment regardless of subjects' level of developmental advancement, but this positive change dissipated within four weeks.
2. The one-sided communication did not have any impact on developmentally advanced subjects neither immediately nor four weeks following the treatment, but did cause a positive shift four weeks following the treatment on non-developmentally advanced subjects.

Interpretations

Based on the above findings the author concludes that:

1. Although one-sided communication is not effective at all with regard to developmentally advanced subjects, non-developmentally advanced individuals seem to benefit from it only after a lag time, but not immediately following the treatment.
2. Attitude dissipation seemingly occurs within three to four weeks following the presentation of a science related persuasive communication built using Shrigley's (1978) design.
3. The demonstrated (short-term) superiority of the two-sided communication suggests a self-persuasion mechanism that may reduce the rate of attitude dissipation associated with attempts of persuasion.
4. Other attitude change models should be examined if long-term attitude change is desirable.

ABSTRACTOR'S ANALYSIS

This investigation addresses two specific questions within the general matrix of studies related to one of the most pervasive and crucial issues in science education (as well as in education in general) namely, "Who says what to whom, when, how, under what particular set of local constraints and with what effect?". The essence of the findings within this particular research setting is two-fold:

(a) two-sided communication is more effective for immediate (but not sustained!) desirable attitude change regardless of the pre-service elementary teachers' level of developmental advancement.

(b) an exposure to one-sided communication brings about a desired change in attitudes four weeks after the treatment (but not immediately after!).

The first finding implies, if generalized, that the attainability of instructionally-based attitude change [which, in turn, is related (as predisposition) to behavioral change (Zoller & Maymon, 1989)] is questionable. The second contradicts most previously reported persuasive communication research. Both may be inappropriately used as an "objective data basis" for decisions and policy-making in contemporary and future science education, particularly as far as the new STS-orientation is concerned.

Consequently, it appears justified to analyze this study on two different levels: the first, its adequacy as far as its internal validity is concerned; namely, the appropriateness of the conceptual framework, methodology, and design within the study per se, the validity of the interpretations and conclusions arrived at, and the consistency of the implications and suggestions derived with the experimental results.

The second level should address the external validity issue. That is, to what extent the conclusions arrived at in this domain specific, limited in scope, study are generalizable and can be extrapolated to the issue of attainability (or feasibility) of preplanned attitude modification (by means of persuasive communication) within the context of science education. After all, attitude and value modification leading, hopefully, to a desirable behavioral change is what education is all about.

As far as the first level is concerned, the research design, methodology, and procedures employed in the study are in accord with well-established experimental practices and the accepted conceptual framework within the matrix of other related studies in the field. However, since within this controlled research design (a) any conclusion arrived at is based on statistical analysis of data obtained from rather small four groups (14 or 15 subjects in each); (b) the target population from which the research sample was drawn is, most probably, local characteristically-bound; and (c) the attitude-related issue in point (energy conservation) is domain specific and contextually bound; (d) the interactions between the integrated conceptual organization and affective framework of instructors and students (the latter being so different from one another to begin with) are complex, multifaceted and unreproducable from one class situation to another, there is not sufficient ground for the extrapolation of the conclusions reached - which are valid intrinsically within this particular investigation - beyond this framework. As stated by the investigator, ..."significant in this study was the finding that the two-sided communication, designed specifically for subjects developmentally advanced (within the domain of energy conservation), was equally effective in changing the attitudes of all subjects regardless the level of developmental advancement immediately following the treatment." This finding corroborates previous related studies suggesting that... "in attempt to induce (immediate) attitude change... the level of advancement [and may I add: the conceptual framework of education and instruction] associated with the communicator via the communication is most critical."

One of our major concerns in contemporary (and future) science education is the development of the critical thinking capability of students defined broadly as ..."a reflective and reasonable thinking that is focused on deciding what to <u>believe</u> or <u>do</u>" (Ennis, 1987). Attitude change is clearly an important factor in attempts to achieve this goal. Consequently, at least from this perspective, two sided (and, preferentially, multi-sided) communications -- which are more relevant and related to higher levels of thinking (i.e., analysis, synthesis and evaluation) and more persuasive with respect to the attitude domain -- should be used and encouraged in science instruction and education rather than the one-sided persuasive messages.

Therefore, the abstractor's view is that although one of the study's findings -- which contradicts most previously reported results -- was that attitudes of non-developmentally advanced subjects show a positive change only after four weeks following the exposure to one-sided communication, the investigator's call for further study of the relationship between domain specific non-developmentally advanced individuals and attitude change as a result of one-sided persuasive communications is <u>not</u> warranted.

It was shown by previous studies (quoted by the investigator) that thoughts generated about communication are more easily remembered than the arguments contained in the communication itself and, therefore, one's own pro- or counter-arguments should be more meaningfully and extensively processed than the arguments that are contained in the communication.

The last statement, I believe, is the most important issue related to the four-part summary question - "who says what with what effect" - and served as the theoretical undergirding for this investigation. Therefore, although one can possibly extrapolate and generalize the results of this study in terms of a valid statement like: learning of adequate persuasive material in an appropriate setting through appropriate means can facilitate persuasion (hopefully in the "right"/desired direction), a significant part of which is attitude modification, further study is needed to facilitate a clearer understanding of the relationships/interactions among instructors' basic educational outlook, teaching styles/methods employed, students' characteristics/learning styles and students' performance/achievements in the cognitive, attitudinal/affective, and behavioral domains.

This is an immense task, the required target-oriented research for which is indeed warranted. Since no one single study can address itself to all the explicit and implicit questions/issues involved, future studies within this matrix of investigations should be designed in such a way that at least one of the meaningful emerging conclusions will have the "generalizability power" - concerning one or more of the above issues - derived from its external validity. This kind of research will guarantee a meaningful contribution towards our capacity to achieve, through the application of appropriate teaching strategies, the desired, agreed upon "high level" goals in the cognitive, affective, and behavioral domains in science education.

REFERENCES

Ennis, R. H. "A Taxonomy of Critical Thinking Dispositions and Abilities." In J. B. Baron & R. J. Sternberg (Eds.), Teaching Thinking Skills: Theory and Practice. New York: W. H. Freeman, pp. 9-26, 1987.

Incso, C. Theories of Attitude Change. New York: Appleton, 1967.

Koballa, T. R. and R. L. Shrigley. "Credibility and Persuasion: A Sociopsychological Approach to Changing the Energy Attitudes of Preservice Science Teachers." Journal of Research in Science Teaching, 20: 683-696, 1983.

Shrigley, R. L. "The Persuasive Communication Model: A Theoretical Approach for Attitude Change in Science Education." Journal of Research in Science Teaching, 15: 335-341, 1978.

Zoller, U. and T. Maymon. "Attitude Modification in Health Education Through an Interventive, Anti-smoking Program Incorporated Within Traditional Science Curriculum." Journal of Research in Science Teaching, 26: 0000, 1989.

Riley, Joseph P. "The Effect of Teachers' Wait-Time and Knowledge Comprehension Questioning on Science Achievement." Journal of Research in Science Teaching, 23 (4): 335-342, 1986.

Descriptors--*Academic Achievement; Elementary Education; *Elementary School Science; *Questioning Techniques; Science Education; *Time Factors Learning

Expanded abstract and analysis prepared especially for I.S.E. by David R. Stevenson, Halifax County-Bedford District School Board, Halifax, Nova Scotia.

Purpose

In an experimental study the author investigated the relationship among the amount of time a teacher waits after asking a question, the cognitive level of teacher questioning, and the effect on student achievement. Three hypotheses were stated:

1. There is no difference in student science achievement related to the effects of cognitive level of teachers' questions.
2. There is no difference in student science achievement related to the effects of wait-time.
3. There are no significant interactions between wait-time and the cognitive level of teachers questions.

Rationale

Both the cognitive level of teachers' questions to students, and the length of time teachers wait before responses are accepted, have been found to produce achievement results, but experimental studies have fallen short of establishing causal links (Riley, 1981). This study sought to extend the research through the variation of wait-time while in association with higher cognitive level questioning, and to establish a relationship between the variables as they affect student achievement.

Research Design and Procedure

Twenty-six preservice teachers in their first intern experience each randomly selected from their internship classrooms five students as subjects. The resulting 129 subjects represented an even distribution across grades 2 to 5.

The preservice teachers were randomly assigned thirty-minute scripted lessons from the Teaching Improvement Kit (Popham, 1972) which contained a specified level of questions and wait-times. The questions had been categorized, using Bloom's Taxonomy (1956), as comprehension, knowledge, or a 50/50 combination of them. A wait-time of 1, 3, or 5 seconds was attempted by asking the subjects to wait until called upon to respond.

The preservice teachers had one hour of practice during which they worked with a peer and audio-taped the practice delivery of the lesson. Treatment lessons to their subjects were then taped and wait-times were measured with a servo-chart recorder. Wait-time was averaged over a whole lesson, and the preservice teachers were classified into groups based on the measurement. The observed average wait-times were 1.17, 3.35 and 5.9 seconds.

The subjects completed a 25-item achievement test at the end of the lesson, with 15 comprehension level questions (as judged by a panel of experts) from the posttest of the Kit and 10 knowledge level items constructed by the author. Both reliability and generalizability were determined and reported.

A 3x3 factorial design, with all factors fixed, was used to analyze the data, with the individual subject as the experimental unit. Analysis of variance and the posttest Newman-Keuls procedure were used to detect differences among group means.

Findings

The author described the results according to the three hypotheses. Hypothesis 1 was rejected on the general finding that the analysis of variance indicated a significant difference related to the cognitive level to teachers' questions, together with the subjects in the 50/50 knowledge/comprehension combination group showing significantly higher achievement than the other groups. Hypothesis 2 was rejected, for the analysis of variance and post hoc analysis indicated significantly higher scores in favour of teachers who used the longest wait-times. Hypothesis 3 was rejected also, for there were significant interactions, with the combination group and longest wait-time showing the strongest results.

Interpretations

The author was able to conclude from the analysis of the comprehension subtest that the use of the 50 percent combination of comprehension and knowledge questions gave strong evidence of a statistical difference. This suggests that teachers are able to improve student acquisition of science comprehension objectives by using such a combination of questions. Further, there was support for the research hypothesis that wait-time influences student achievement, with wait-time beyond a supposed threshhold leading to higher science achievement. Interpretation of the interaction occurring on the knowledge test suggested that extended wait-time may be inappropriate for knowledge or low level questions, and that a wait-time threshhold may exist.

In summary, the results of the study indicated that, for achievement of comprehension level objectives, a combination of low and high cognitive level questions together with long wait-times seemed more effective than the other tested combinations.

ABSTRACTOR'S ANALYSIS

Two topics which have entered the general discussion in school staff rooms have been the cognitive (and other) categories of educational objectives, and the idea of wait time. The extent to which either is taught effectively to preservice teachers and to teachers now in service will have to be discussed elsewhere. The author of the study under review has shown that attention to the topics can have effects.

Riley is exploring a fruitful avenue for teachers in this investigation. With demands on teachers in their classrooms increasing, any measure which may improve on effectiveness while retaining humanity is to be considered. The possibility of differentiating among levels of wait-time and of levels of questioning seems promising in general.

The experimental setting for the study has a strength in that it has permitted close attention to the selected variables in a way which is difficult for teachers in a classroom with 25 to 35 students. Thus, the gains which have been found to be possible in the group of five subjects may need further consideration, for the demands upon teacher attention in the whole-class setting may cause an erosion of the effects found under the conditions described for the study. It would be helpful, for example, to have results for the same preservice teachers who might have been given the opportunity to present useful lessons to whole classes from which the subjects were chosen, with the preservice teachers continuing the combination of cognitive levels and wait-times.

Additional information may help the reader understand better the significance of the author's findings. While the random assignment of students to treatments was in turn followed by the random selection of subjects, it could be helpful to know some descriptive features of the administration of the schools and some of the community characteristics which affect the teachers and their students. Was the study undertaken in schools which are research "smart"? Could a bias toward streaming unintentionally have affected the outcome?

Science in elementary schools, in concert with the developing understanding about language and mathematical concept acquisition, requires a strong tie to the real world through meaningful activities. The science which is usually advocated is exploratory in nature, consisting of investigations whenever possible, preceded by and followed by questioning and discussion and personalized student note-taking.

The study adds knowledge to our understanding of wait-time and levels of questioning, but one is urged to consider the generalizability of the results. The topic of wait-time has been explored for some time, and the results seem clear. The same clarity may not be voiced for the cognitive level of teachers' questioning, for the variables have eluded concise investigation. That situation may still exist, but the gap seems to be narrowing as a result of this experimental study.

REFERENCES

Bloom, B. S., M. D. Engelhart, E. J. Furst, W. H. Hill, and D. R. Krathwohl. (Eds.) "Taxonomy of Educational Objectives: The Classification of Educational Goals." Handbook I: Cognitive Domain. New York: David McKay, 1956.

Popham, W. J., E. Baker, J. Millman, and J. D. McNeil. Teaching Improvement Kit. Los Angeles: Instructional Appraisal Services, 1972.

Riley, J. P. "The Effects of Preservice Teacher's Cognitive Questioning Level and Redirecting on Student Science Achievement." Journal of Research in Science Teaching. 18 (4): 303-309, 1981.

Okebukola, P. A., and M. B. Ogunniyi. "Effects of Teachers' Verbal Exposition on Students' Level of Class Participation and Achievement in Biology." Science Education, 70: 45-51, 1986.

Descriptors--*Academic Achievement; *Biology; High Schools; Science Education; *Secondary School Science; *Student Participation; *Teacher Behavior; *Verbal Communication

Expanded abstract and analysis prepared especially for I.S.E. by David L. Haury, Tufts University.

Purpose

The purpose of this study was to compare the effects of two instructional strategies on student achievement and participation in biology classrooms. The two strategies differed in the nature of instructional discourse, with the effects of a direct approach to instructional discourse being compared to those of an indirect approach. A teacher taking the direct approach to verbal interaction with students is characterized as one who lectures, gives opinions, and presents overviews of material. Conversely, the teacher who takes an indirect approach accepts feelings and student ideas, relates expressed feeling to probable causes, and praises student contributions.

Two hypotheses were examined. One hypothesis was that the two instructional strategies have differential effects on student achievement in biology. The second hypothesis was that the two strategies also have differential effects on the level of student participation in class. Since conflicting results have been reported from previous studies of instructional discourse, there were no stated expectations about which instructional approach, direct discourse or indirect discourse, would foster greater achievement or classroom participation.

Rationale

The study reflects a concern about how to optimize instructional effects in terms of student achievement and classroom participation, given the constraints of high student to teacher ratios and a lack of laboratory facilities for science. There is an implicit assumption that a laboratory-based approach to science instruction would be desirable and most effective, but circumstances prevent such an approach. With inadequate facilities and large classes, then, the need is to promote the most effective form of teacher discourse.

The measure used to compare instructional approaches is derived from interaction analysis, and is expressed in the form of a ratio of "indirect" to "direct" teacher-student interactions. Mixed results have been reported from studies comparing direct versus indirect forms of discourse, so this study is an attempt to bring some resolution to the issue in terms of student performance in biology classrooms.

Research Design and Procedure

A quasi-experimental research strategy was employed, with instructional treatments being randomly assigned to fifteen class groups comprising 750 high school students. Using Campbell-Stanley notation, the nonequivalent control group design is represented as follows:

O_1	X_1	O_2	O_3	
O_4	X_2	O_5	O_6	
O_7		O_8	O_9	

Key

X_1 = Direct Instructional Discourse

X_2 = Indirect Instructional Discourse

O_1, O_4, O_7 = BAT Pretest

O_2, O_5, O_8 = CPS Inventory

O_3, O_6, O_9 = BAT Posttest

Fifteen randomly selected preservice biology teachers were randomly assigned to the fifteen classroom groups of students, with ten of the teachers trained in one of two forms of instructional discourse. Five teachers received training in a direct form of instructional discourse, and five received training in an indirect form of instructional discourse. Teachers of the five comparison, or control, groups were not assigned to a special training program.

Instruments used were the Flanders Interaction Analysis Category System (FIACS) to measure the level of direct and indirect instructional discourse, a Biology Achievement Test (BAT) to measure student achievement, and a Class Participation Scale (CPS) to measure student participation in class. The FIACS is a 10-category classroom observation instrument frequently used in research focusing on analysis of classroom interaction. The BAT is a 40-item multiple-choice test apparently created by the authors, and the CPS is a 10-item observation inventory developed in an earlier study. Acceptable reliabilities were reported for the BAT and CPS.

The preservice teachers who provided the experimental instruction were exposed to a structured training program consisting of two lectures, four demonstrations, two practice lessons, and an undisclosed number of discussion sessions. The experimental instruction was then implemented over a two week treatment period, with four lessons on photosynthesis being presented each week, for a total of eight lessons. The BAT was administered as both a pretest and posttest, and a five-member team of observers employed the FIACS and CPS to collect classroom data during the instructional treatments. The observers learned to use the FIACS and CPS through use of audiotapes and trial lessons, and FIACS scores were used following treatment to verify the instructional behavior of teachers.

Analysis of variance was used to check for differences in pretest scores, posttest score, and level of class participation among student groups. Descriptive statistics were used to characterize the teacher-student interactions of each group.

Findings

Ratios of indirect to direct teacher-student interaction were calculated for each treatment group, with those trained in direct instructional discourse exhibiting the range of lowest values (0.15-0.22), those trained in indirect instructional discourse exhibiting the range of highest values (1.83-1.93), and those receiving no special training exhibiting a broad range of intermediate values (0.36-1.51). Analysis of variance revealed no significant differences between groups on the pretest, so the treatment groups were considered to be initially equivalent in terms of biology achievement.

Analysis of variance did reveal statistically significant differences, however, between the two experimental groups with regard to posttest performance and level of classroom participation. The treatment group exposed to the indirect form of instructional discourse obtained a higher mean score on the posttest and a higher mean score on the CPS. It is implied that a higher CPS score represents a higher level of classroom participation.

Interpretations

The results are interpreted to mean that indirect instructional discourse fosters higher levels of achievement and classroom participation in biology than does direct instructional discourse. In discussing this finding, the authors surmise that an indirect approach has such an effect because students are more relaxed and free to interact productively with teachers and materials. Mention is made of the purported relationship between a relaxed classroom atmosphere and student creativity. A parallel is then drawn between an indirect instructional style and science, an enterprise which involves curiosity, development of a searching mind, and the sharing of opinions.

Given the limitations of the study, the authors close by suggesting areas where further research is needed. They suggest that the influence of instructional discourse on student creativity be examined, they mention the need to study the effects of student gender, background, and ability on the level of classroom participation, and they suggest that other dimensions of instructional discourse be examined for effects on student performance.

ABSTRACTOR'S ANALYSIS

The study was a technically well-conceived examination of an important issue, and it provides potentially useful information about the predominant form of instruction in many science classrooms: teacher discourse. Some ambiguities about the treatments lead to questions about the findings, but the article clearly demonstrates that instructional discourse can take a variety of observable forms which have differential effects on student learning. The findings are congruent with evidence that active engagement fosters learning, if one considers indirect instructional discourse a strategy that invites cognitive involvement among students. This is a rather intuitive idea, but one gains confidence in intuitive ideas when they are contextualized and tested in natural settings. The investigation reported in this article is exemplary in having embedded a solid research design within a natural setting in seeking answers to straightforward questions.

Unfortunately, it is somewhat difficult to generalize from the stated findings, and the difficulty arises from ambiguities about the differences among instructional treatments and their measurement. Within the introduction to the article, there is some ambiguity about the actual differences between direct instructional discourse and indirect instructional discourse, with some of the same words being used to describe both approaches (i.e., asks questions, gives directions, gives facts or gives factual information). Most educators have an intuitive notion about the differences between direct and indirect instructional discourse, but the differences have to be stated in very clear, operational terms to aid in the application of findings or further examination of hypotheses. The problem would have been resolved if the training program had been described in more detail, or if the FIACS had been briefly described. As it is, one must refer to the original source of the FIACS to interpret the reported ratios which are measures of "indirectness" in instructional discourse. Furthermore, because there is no statistical test of differences among the ratios, there is uncertainty about whether the reported differences in instructional discourse among teachers are significant. It must be acknowledged, however, that the authors took great care to randomize selections and assignments as much as possible, and they checked inter-observer reliability to reduce threats to validity.

Ambiguities about the BAT and the CPS also give rise to open questions about the effects of the instructional treatments. Readers are told very little about the BAT. Is it a general test of knowledge in biology, or is it tailored to instruction about photosynthesis? How was validity judged, and who made the judgments? Providing sample test items would have been valuable in this instance; they would allow the reader an opportunity to judge face validity. Sample items were provided for the CPS, but the reader is not told how the items are scored. And what is meant by classroom participation? Does the CPS measure the degree of active involvement, or does the scale measure variety in the forms of involvement? These questions do not prevent the reader from acknowledging effects in student performance due to treatment, but they do make it a little difficult to generalize from the results. It is also difficult to judge the educational significance of the findings.

There is also a question about the analysis of variance. It is reported that one-way analyses of variance were computed, but there were two dependent variables and three treatment groups: two experimental treatments and one comparison group. Such a design calls for a form of multivariate analysis that can simultaneously accommodate both the effects across several groups and the effects of more than one dependent variable. Very little is reported about the instructional treatment of the comparison group, and the authors do not report whether there were statistically significant differences in achievement and class participation between the comparison group and the two experimental groups. The comparison group is used, in effect, only to illustrate that training enables teachers to employ a more direct or indirect style of instructional discourse as measured by the FIACS; BAT and CPS scores for the comparison group are not employed in testing the significance of differences in effects among groups. Given the intriguing result that the mean BAT score for the control group was even lower than the score for the group exposed to direct instructional discourse, a test of significance would have been useful. Given the large sample size and number of classroom groups, descriptive data about the individual groups would also have aided interpretation of the results. In fact, the experiment was replicated five times in that there were five discrete groups for each instructional treatment, so multiple comparisons across groups may have yielded a more compelling demonstration of the effects due to treatment.

In the final section of the article, labeled "Discussion", the authors did not actually discuss the findings at length or bring closure to the investigation. Rather than refer to issues raised in the introduction or interpret the findings for the reader, the authors speculated about the parallels between scientific enterprise and the indirect verbal influence of teachers. Some cultural comparisons were also made between African countries and the developed countries of North America and Europe, and the relationship between classroom climate and creativity was mentioned. As interesting as these matters may be, the authors really did not take the opportunity to help readers interpret the findings in terms of the original concerns, and they did not provide any direction for those who may be interested in either applying the results or following up with related studies. One is left with the sense that a very important study was conducted, but readers are given only a tantalizing glimpse of the outcomes.

Ogunniyi, M. B. "An Investigation of the Nature of Verbal Behaviors in Science Lessons." Science Education, 68 (5): 595-601, 1984.

Descriptors--Biology; Chemistry; Interaction; Physics; *Questioning Techniques; Science Education; *Science Instruction; Science Teachers; *Secondary School Science; *Teacher Behavior; Teacher Student Relationship; Time Factors Learning; *Verbal Communication

Expanded abstract and analysis prepared especially for I.S.E. by Gene F. Craven, Oregon State University.

Purpose

The stated purpose of this study was "to identify and describe the nature of teachers' verbal behaviors exhibited during science instruction. No attempt was made to isolate possible effects of underlying conditions such as teacher's sex, age, amount of teaching experience, adequacy of learning, materials and aids, etc." (Ogunniyi, 1984, p. 595)

Rationale

Concern about poor science performance of students on West African School Certificate Examinations is reported to have resulted in a series of studies showing that "science teaching in Nigeria has been adversely affected by the lack of adequately trained personnel and facilities." The present study examines the type of verbal interactions that occur during science instruction. The investigator assumes that "the quality of verbal exposition occurring in science lessons will affect a student's understanding of the given subject matter." (Ogunniyi, 1984, p.595)

Ned Flanders (1970) has been a major contributor to research on teachers' verbal behavior and on classroom interaction analysis. Ogunniyi and many other researchers have used modifications of the Flanders' technique to investigate teachers' verbal behaviors and their effects on student learning.

Flanders' purpose for interaction analysis was to study teaching behavior by keeping track of selected events that occur during classroom interactions. He proposed two applications of data obtained via classroom interaction analysis studies as first, helping the individual develop and control his or her teaching behavior, and, second, to discover through research how to explain the variations which occur in the chain of classroom events (Flanders, 1970).

Research Design and Procedure

Twenty-four randomly selected secondary school science teachers (about 20% of the total population) from the Ibadan Local Government area of Oyo State, Nigeria, were observed in this study. Data are reported for eight teachers in each of three subject areas: biology, physics, and chemistry. Each teacher taught an average of 30 students per class. Each was observed for an entire class period two or more times within a one-month period.

Three observers were trained to use a slight modification of Flander's Interaction Analysis Categories. Teacher questions (category 4) were divided into additional subcategories; memory (m), informational (i), rhetorical (r), leading (l), and probing (p).

Training of the observers occurred in three stages: (1) one week of explaining the investigation, memorizing the interaction categories and observing a series of contrived science lessons given by the investigator; (2) one week for practicing how to identify and categorize verbal behaviors in different classrooms with all the observers staying in a class at the same time and comparing notes after each lesson; and (3) intensive practice in classroom observation.

Inter-coder agreement percentage among the three trained observers was reported to range between 85 and 95%. Before actual classroom observations began, agreement of specific question identification and categorization among the three observers is reported to have been between 95 and 98%. Average inter-rater reliability correlation was reported as 0.88.

To enhance the reliability of recorded observations, a set of procedures was established for use by each observer to (1) establish a warm rapport with teachers and their students; (2) record every two seconds; (3) use an audiotape recorder to record the entire period; (4) avoid any form of distraction during the observation period. Also, (5) post observation comparison of notes and tape recordings with actual observations, (6) identifying each specific classroom behavior as closely as possible, and (7) recording only behaviors with the greater "liklihood of occurrence" in cases of ambiguity or when two events occurred simultaneously. A correction formula was used to correct for chance guessing that may have resulted from the two-second categorization schedule while collecting data.

Data are presented in four tables: (I) percentage of responses for each of the eleven categories are reported for lessons 1 and 2 for the eight teachers in each subject area; (II) percentage of verbal interactions for teachers (sum of categories 1-7) and students (categories 8-11) for each of the two lessons in each of the three subject areas; (III) percentage of subcategory types of questions used by teachers in science lessons in each of the three subject areas; and (IV) rate of question asking by teachers and by students in each subject area near the beginning, the middle, and the end of the lesson period.

Findings

Lessons taught by the 24 teachers who participated in the study were dominated by the teacher, with most of the students remaining passive. More than four-fifths of the lesson period was devoted to verbal instruction and about one-fifth of the class period devoted to giving directions, mainly concerning the instructional process. Teachers tended toward supplying their students with facts "which they absorbed verbatim."

In spite of current emphasis on inquiry lessons, science teaching in the schools observed was "mainly a talk-and-chalk affair." Students' verbal expressions were limited mainly to responding to teachers' questions. If the learner asked any question at all, it was to seek information. The greater the number of teachers' questions, the greater the number of students' responses. Conversely, the greater the frequency of criticism, the lesser the frequency of students' self-initiated talk. Except for biology classes in which it was slightly higher, silence or irrelevant talk accounted for less than 5% of the class period.

A majority of the teachers spent more than one-tenth of the class period asking questions. Most were factual questions with biology teachers asking more informational and memory questions than did chemistry or physics teachers. The latter two groups asked more rhetorical questions which tended to generate chorus responses. Teachers in all three subject areas asked relatively low percentages of leading and probing questions.

Apart from biology teachers who had a lower rate of questioning, chemistry and physics teachers asked an average of 1.51 questions per minute. This was about three times the rate at which students asked questions.

Interpretations

It is evident, the investigator concluded, that science teachers involved in this study spend a considerable amount of time on verbal instruction. Their students, on the other hand, passively listen and absorb or record information verbatim. He states that "it is quite doubtful that students who merely listen, absorb, or record information verbatim develop the high level of cognitive thought expected of the new science curricula."

The teachers involved in this study were found to ask more questions than their students. "In one sense," the investigator writes, "this indicates that the students are unable to answer all the questions posed by their teachers. In another sense, it merely confirms the status quo in a traditional society where the elder (in this case the teacher) has all of the necessary knowledge."

While reporting that most of the teachers exhibited identical verbal behaviors across the three science disciplines, the investigator concluded that biology and chemistry teachers tended to reward and reinforce their students' responses more than did the physics teachers. Unlike their teachers, physics and chemistry students tended to maintain a "low profile" throughout the lesson period.

"The fact that all three groups asked relatively lower percentage of leading and probing questions ... may be related", the investigator writes, "to the information-based examination." "It would seem", Ogunniyi concludes, "that students perceive teachers' questions as a form of criticism, authority, or accountability which puts some limitations on their personal questions or self-initiated talk. In such a situation, all the students can do is to answer teachers' questions while posing very few of their own questions. In a traditional society such as Nigeria, this is not difficult to understand."

ABSTRACTOR'S ANALYSIS

Relationship to the Matrix of Studies in the Area of Research - Investigations of effects of teachers' verbal behaviors were pioneered by Ned Flanders with five studies conducted betweeen 1959 and 1967. A ten-category system (Flanders' Interaction Analysis) was developed to test the hypothesis that teacher "indirectness" has a positive effect on student attitudes and subject matter achievement. More than a hundred studies using the Flanders' Technique were eventually published (Brophy & Good. 1986). Research data generally support Flanders' hypothesis of a positive relationship between "indirect" teaching and attitude (liking for the teacher and class). Several experimental studies comparing indirect to direct teaching failed, however, to show a significant relationship between "directness" or "indirectness" and group achievement (Rosenshine, 1970). Thus, the present study builds on a rather extensive research base.

Ogunniyi used a "slight modification" of Flanders' Interaction Analysis Categories and a rather significant modification of the Flanders' Technique: that being the absence of preassessments of attitudes and achievement. In this abstractor's judgment, the present study makes no new conceptual or methodological contribution to either the Flanders' Technique or to research on teachers' verbal behaviors or their effects on "the science performance of students." And, the researcher is not specific as to how his research-based description of "the nature of teachers' verbal behaviors exhibited during science instruction" is to improve science education in Nigeria.

Validity of the study. To what extent are the investigator's interpretations of the data useful in identifying and describing science teachers' verbal behaviors that may relate to "poor science performance of Nigerian students on the West African School Certificate Examination?" They are useful in identifying and describing teachers' verbal behavior only. With a preassessment of achievement, it may have been possible to have found some relationship between teachers' verbal behavior and "science performance of Nigerian students."

The Flanders' Technique is generally considered to produce valid data about verbal behaviors during group instruction. Advantages of the Flanders

Technique are that (a) it has minimal effect on tne regular curriculum, thereby facilitating access to classrooms for observational inquiry; (b) it is easy to learn; and (c) inexpensive data may be generated using the system.

Ogunniyi's version of the Flanders' Technique was appropriate for his purpose of identifying teachers' verbal behaviors and a quantitive description of the nature of teachers' verbal behavior during science instruction. Without an achievement preassessment, the data do not permit a test of his assumption (hypothesis) that "the quality of a verbal exposition occurring in science lessons affects students' understanding of the given subject matter," but this was not a stated purpose of the study.

Research design. The present study is best classified as a quantified status study. It meets criteria for neither experimental nor quasi-experimental designs for research on teaching (Campbell & Stanley, 1963). Nor does it meet a basic criterion of qualitative research in that the researcher had identified definite verbal behaviors to look for prior to entering the classroom to observe (Taylor & Bogdan, 1984). Qualitative research strategies would have been preferable to the Flanders' Technique to validly "describe the nature of teachers' verbal behavior exhibited during science instruction" in that they would have centered on what actually occurred in those classrooms rather than categories of behaviors that were established for a somewhat different purpose.

It would be useful to know more about the qualifications of the persons who were trained as observers; if the researcher was one of those observers; and in which subject matter specialty or specialties each observed. These are factors that could have a bearing on the reliability and hence validity of data that are reported.

Appropriate procedures were followed in training the observers. Pre-observation inter-coder agreement (85-98%) and inter-rater reliability correlation (0.88) values are acceptable. Use of the audiotape recorder permits reexamination of the verbal interactions to check on accuracy of the on-line data collection and to confirm inter-coder agreement and inter-rater reliability after making the classroom observations. Doing so would provide some guarantee against observer skill deterioration during the study period. On-line performance of coders in the science classroom may differ from that of coding samples in a laboratory or training session.

Reliability Estimates Which Relate to Validity of the Data. High (85-98%) inter-observer agreement may or may not mean agreement with what actually occurred because systematic misinterpretation can exist even with high agreement. To check on the possibility of misinterpretations, it would have been desirable to have compared observer scores with a criterion, e.g., scores of an "expert" observer, using data on randomly selected audiotapes of classroom instruction.

Also, reliability may suffer when observers are forced to make judgments every two seconds in complex classroom settings where mixed verbal and nonverbal interactions (procedural, behavioral, academic) are occurring in rapid order. It would be useful to know the extent to which the audiotapes were used to reconfirm in-class coding by listening to the verbal interactions in a more relaxed setting where it is possible to rerun the tape. The investigator does not state whether the reported data are on-line, in-class data or data that have been verified via an examination of audiotapes. The research procedure, "post-observation comparing of notes and comparing tape recordings with actual observations," appears to imply a reexamination of the on-line data.

Random selection of teachers is appropriate for a quantitative study and 20% of the entire population is an adequate sample for the purposes of this study. Randomization is assumed to result in "equivalent" groups of subjects. While this study excluded teacher and school factors, summaries of subject area factors such as school size, community type, education level of teacher, and length of teaching experience would help to assure the reader of school and teacher equivalence across subject areas.

Observers' initial impressions and expectations are known to have a distorting effect on later judgments. And influence of the observer on the classroom can not be overlooked nor can expectations and reactions of the persons being observed. It would be useful for the reader to have information concerning initial school/teacher contacts; how the purpose of the investigation was presented to the teachers; how rapport with teacher and students was established and how the observer's effect on classroom interactions was minimized. Good research procedure would include desensitizing students and teacher to the presence of the observer prior to collecting research data. Yet, the report makes no reference to observer presence in the classroom prior to the two observations to collect data. It would be desirable to know how the actual lessons to be observed were selected and the average time interval between observations of teachers in each subject area.

The simple percentage statistic used in analyzing and reporting the data is appropriate for the purpose of the study, i.e., to identify and (quantitatively) describe teachers' verbal behavior during science instruction.

Adequacy of the written report. Journal editorial policies place many restrictions on the writer of any research article. It is difficult to provide all of the information desired by any reviewer in a six-page journal article. Given the page restrictions placed on the investigator, the report is written clearly and is adequate for most readers. Critical reviewers of the study would welcome a more detailed report as can be inferred from this analysis.

The number of published studies using the Flanders' Technique has declined significantly during the past decade. Only two studies using the Flanders' Technique in science classrooms appear in ERIC abstracts in the last five years; the current study and one by Musoko in Kenya.

In reporting his findings, Ogunniyi states rather frequently that his finding is to be expected. He cites several of his findings as evidence which "confirms findings in other parts of the world." If nothing new or unexpected is being discovered, there does not appear to be a need for more confirmational studies. Ogunniyi does provide some quantitative generalizations about verbal behaviors of Nigerian science teachers but he does not share with the reader how he perceives using that information to improve science education in Nigeria.

Additional studies using the Flanders' Technique may be justified if the intent is to sensitize teachers to classroom verbal behaviors (one of Flanders' goals). And, it may be desirable to provide other quantitative descriptions of verbal behaviors of science teachings in Nigeria, Kenya, and other countries if those descriptions can be logically shown to "improve science performance of students."

Descriptive research appears preferable to quantative research when the purpose is description of behaviors in the classroom. Descriptive researchers, observing everyday classroom situations over an extended period of time could obtain data that might be used to develop a classification scheme different from that of Flanders. Such a scheme might result in a "new" area of research on teachers' verbal behaviors and their effects on desired outcomes of science instruction.

REFERENCES

Brophy, Jere and Thomas L. Good. Ch. 12. "Teacher Behavior and Student Achievement", in Wittrock, M. C., editor, Handbook of Research on Teaching, 3rd Ed., Macmillan Publishing Co., 1986.

Campbell, Donald T. and Julian C. Stanley. Ch. 5. "Experimental and Quasi-Experimental Designs for Research on Teaching", in Gage, N. L., Editor, Handbook of Research on Teaching. Rand McNally & Company, Chicago, 1963.

Flanders, Ned. A. Analyzing Teaching Behavior. Addison-Wesley Publishing Company, 1970.

Rosenshine, B. "Experimental Classroom Studies of Indirect Teaching." Classroom Interaction Newsletter, 5 (2): 7-11, 1970.

Taylor, Steven J. and Robert Bogdan. Introduction to Qualitative Research Methods: The Search for Meanings, 2nd Ed. John Wiley & Sons, New York, N. Y., 1984.

Hudson, H. T. "A Comparison of Cognitive Skills Between Completes and Dropouts in a College Physics Course." Journal of Research in Science Teaching, 23(1): 41-50, 1986.

Descriptors-*Academic Achievement; *Cognitive Processes; *College Science; *Dropouts; Higher Education; Mathematics Achievement; *Mathematics Skills; *Physics; Problem Solving; Science Education

Expanded abstract and analysis prepared especially for I.S.E. by Ubiratan D'Ambrosio, Interdisciplinary Center for the Improvement of Science Education, Brazil.

Purpose

This study was designed to determine if the act of providing diagnostic feedback to students had any impact in the correlation between performance in physics and a test of simple, mechanistic mathematics skills, and at the same time to determine if students who drop out of the physics course demonstrated any identifiable difference in performance on tests involving a variety of reasoning and problem solving skills.

Rationale

One would expect that lack of mathematical skills would be a major factor in dropping out from physics at the collegiate level. Previous research has given less than conclusive views on this question. The author himself had shown that precourse tests of mathematical skills did not serve as a discriminator between students who would ultimately complete the course and those who dropped it. Further research introduced together with mathematical skills items requiring mathematical reasoning with general indicators of past performance and, again, the scores on mathematics tests by future completes and dropouts were not so separated to be useful as predictors. The measures did not correlate with the performance of the dropouts at the point of dropping, nor predict whether or not an individual is at high risk of dropping the course.

One would expect that poor performance in algebra and trigonometry would be an indicator of future poor performance in physics, and would put some individuals in the high risk category for dropping out. The same argument goes for problem-solving skills and reasoning ability. Research results go against the conviction that mathematical skills help for a good performance in physics. One explanation is that during the first classes of physics, a growth in interest together with the reviewing processes and homework of a simple mathematical nature and closely related to the subject matter helps in overcoming the lack of mathematical skills and thus determine, better than did previous training and knowledge, performance in physics. This is indeed a special form of diagnostic/remedial instruction. In fact, many places provide workbooks to students after their tests and thus the scores of the test are lessened as indicators of future performance.

So, it is desirable to know whether or not the correlation between scores on a mathematical skills test and physics performance would be dramatically affected by the simple act of making the results of the diagnostic testing available to the students, and providing them with review material.

Research Design and Procedures

The author designed three instruments administered to 152 students in the first semester General Physics course, in the first week. Forty percent of the students had completed calculus and 56% had no previous high school or college physics, although only 16% were freshmen.

The first instrument was a 31-question test of mathematical skills developed by the author together with R. Rottman (Journal of Research in Science Teaching, 18: 291-294, 1981) with three additional questions. The second was an eight question test of proportions, involving both direct and inverse ratios. The third instrument was an eleven question test involving general word problems requiring translation of words into symbols and symbols into words. The paper gives, as appendices, sample questions of both second and third instruments. All the questions are related to physics and contain topics usually required to solve physics problems. The questions were constructed in consultation with physics faculty and relied upon current literature on similar research.

Two dependent variables were used in the study, the first major course examination, given approximately one month after the instruments were applied, and a repeat of that examination. The final physics grade for the course consisted of the sum of three major course examinations, a comprehensive final examination, daily homework, and a weekly free-response problem solved in class.

After excluding all the students who did not submit to the three instruments and both the first major course examination and the repeat of it, 87 subjects were left, 64 who completed the course and 23 who dropped.

Findings

The paper gives three tables of scores and their analyses.

Some of the findings say that dropouts scored numerically lower than completes on all precourse measures, but the mean scores are not so different to provide a discriminatory indice. For the completes the correlation between the mathematics score and both proportions and translations is not significant, in contrast with the results for dropouts. For dropouts, only translations and the sum (proportions plus translations) correlated at $p < 0.05$ with the first major exam, and for completes all precourse measures correlated with physics grade at $p < 0.01$. The correlation with final grade is 0.24 for mathematics, 0.49 for the first major exam and 0.62 for the repeat.

Clearly, there is a possibility of finding significance by purely random chance. It is interesting to note that while completes and dropouts perform at very close levels on tests of mathematics, proportions and translations, they apparently do not use skills in similar ways when solving physics problems. This leads to the conjecture that completes and dropouts differ in their ability to apply different cognitions to different situations. According to the author, "It is as if the completes possess two independent mental processes while, for the drops, the same mental processes are overlapping and contaminate each other" (p. 47).

The original hypothesis which led to this study seems confirmed. If students are advised on their areas of weakness and self-help material is provided, the correlations between math and physics are lower than when the test is given without feedback. Hence, remediation can be accomplished for a number of students by simply identifying their areas of weakness and making self-study materials available to them.

Interpretations

Reviewing and comparing with results in early research on the correlations between mathematics and physics performance, it seems that, at least for this sample, the impact of mathematics on physics has decreased. Students are able to make-up for deficiencies when they are aware of these deficiencies and are provided self-help material. It is clear that the sample is somewhat atypical of the student population. The research design excluded a significant part of the cases which offer much interest. This is mainly due to the exclusion of those who did not take both the first major exam and the repeat, which are either the best students or those who gave up even before trying it again. Those who remained for the repeat probably received extra help from the instructor, and this extra help, for this kind of population, produces a high payoff.

The results reinforces the impression that pretesting to identify high risk drop-outs is far more complex than some may think.

ABSTRACTOR'S ANALYSIS

This is indeed a very thought provoking paper. Although the research design may leave the reader sometimes uncomfortable, as even the author is, as he points out several times, one does not see an easy way of improving it. The subject is too complex. How can one predict behavior? How can one determine the capability of any individual to retrieve, when conveniently motivated, information and skills that were lying dormant in the mind, hence unidentifiable by tests? How can one say that being able to provide an answer for a question in a certain ambiance will allow the individual to use the skills shown in this case to deal with different situations? These are basic questions in a process largely unknown to us.

It is regrettable that the complexity of the problem and the somewhat clear picture brought to us by results like the one in this paper do not have impact on curriculum planners at the mathematics collegiate level who still insist in "equipping" students to pursue further studies in the sciences and engineering by providing them skills and drilling in calculus and in the so-called, still present although sometimes as remedial, college mathematics. A new approach to the concept of providing basic mathematics which indeed may help students in further studies of science and engineering is absolutely urgent. It is about time we go away from mere drill and skill to a conceptual approach to calculus.

Further research such as the one under review, with more sophisticated design, is absolutely needed and should be better known to mathematics educators at the collegiate level. Mere performance in mathematics is far from an indicator of successful completion of a future scientific career. This surely has to be taken into account by college mathematics curriculum planners.

Johnson, Virginia A. and David Lockard. "The Effects of Kinetic Structure and Micrograph Content on Achievement in Reading Micrographs by College Biology Students." Journal of Research in Science Teaching, 22 (8): 713-721, 1985.

Descriptors--Achievement; *Biology; *College Science; College Students; Higher Education; Photographs; Science Education; *Science Materials; *Visual Aids; *Visual Discrimination

Expanded abstract and analysis prepared especially for I.S.E. by Joel J. Mintzes, University of North Carolina.

Purpose

This study examined the effects of kinetic structure and micrograph content on the acquisition and retention of concepts and skills required to interpret photomicrographs in the biological sciences. Specifically, the investigators studied the effects of instructional units possessing "high" and "low" kinetic structure, and "unified" and "varied" content on the ability to read micrographs depicting structural features of protozoans.

Rationale

The study is grounded in O. Roger Anderson's work (1969, 1971, 1974) on the "kinetic structure" of verbal discourse in instruction. The general proposition advanced by Anderson is that, "acquisition of verbal material is enhanced when contiguous verbal units (statements) in a communication contain identical verbal elements ...." Consequently, variation in student achievement depends, in part, on the extent of "relatedness" or "commonality" found in consecutive elements. (Anderson's measure of commonality, which he calls the "B_1 coefficient," ranges from 0 to 1.0 and represents the average number of shared verbal elements in a set of "discourse units.")

With this effort the present authors sought to further test the applicability of Anderson's theory "within a milieu of visual information." They also suggest that the study is important because "knowledge retention is an aspect of learning that has been neglected in kinetic structure research."

Research Design and Procedure

One hundred students who enrolled in an introductory biology course at Towson State University attended three audiovisual presentations (total instructional time = 48 minutes) and three "micrograph practice sessions" which focused on concepts and skills in the reading of light, transmission electron, and scanning electron micrographs. Students were randomly assigned to one of four treatment groups:

-Low kinetic structure ($B_1 < 0.25$)/varied micrograph content
-Low kinetic structure/unified micrograph content
-High kinetic structure ($B_1 > .45$)/varied micrograph content
-High kinetic structure/unified micrograph content

Students assigned to the varied content treatment groups observed representative micrographs of five different protozoan species; those attending the unified content presentations observed micrographs of a single protozoan species (Tetrahymena). The authors state that the "comparability" of the micrograph sets was evaluated by a panel of judges selected by a "noted microbiologist." In all other pertinent respects the groups were treated identically.

Following the treatments, the Reading Micrographs Skills Test was administered at two intervals: 48 hours after treatment (knowledge acquisition) and three or more weeks after treatment (knowledge retention). The Test was comprised of two subscales labelled G (five short-answer items measuring "general concepts" of micrograph reading; alpha = 0.66) and RM (unspecified number of multiple-choice and short-answer items "related to reading the given micrograph"; alpha = 0.75).

Means and standard deviations were presented in tabular form and separate repeated measures analyses of variance were reported for the G and RM subscale data.

Findings

Results of the repeated measures analyses suggest:

1. Differences on the G (general concepts) subscale as a result of variation in kinetic structure. Students receiving the high kinetic structure treatment scored significantly higher than those in the low kinetic structure treatment.
2. The effects of kinetic structure on G subscale scores were not diminished by time. " ... students receiving high kinetic structure instruction not only learned more ... but they also retained more."
3. No differences on the G subscale as a result of variation in micrograph content and no interaction effects.
4. No differences on the RM (micrograph reading) subscale as a result of variation in kinetic structure and no interaction effects.
5. Differences on the RM subscale as a result of variation in micrograph content. Students receiving the unified treatment scored significantly higher than those receiving the varied treatment.
6. The effects of micrograph content on RM subscale scores were diminished by time. " ... forgetting occurred on Test RM."

Interpretations

The authors drew several conclusions from the study, including the following:

1. "High kinetic structure instruction is more effective than low kinetic structure instruction in developing general concepts about micrographs."
2. "However, no such application of kinetic theory to perceptual learning [micrograph reading] is indicated in this study."
3. " ... a unified visual content may be more effective than a varied content [for enhancing] the organization and transfer of actual reading micrograph skills."
4. "This study establishes that there is a need for instruction in reading micrographs."

ABSTRACTOR'S ANALYSIS

Relationship to other studies. This study is one of a large number of investigations that has grown out of Anderson's work and contributed to our understanding of the importance of structure in verbal communication in the science classroom. In presenting their work, the authors do a good job of summarizing the critical elements of kinetic theory which bear on the questions they pose.

Many biologists would agree that reading and interpreting micrographs is an important skill; one which is poorly addressed in introductory courses. Clearly this is an area that has received too little attention as reflected in the paucity of previous research (Ali, 1984; Montgomery, 1979).

One issue that remains a bit puzzling, however, is the choice of kinetic theory as a framework for analyzing a learning task which is essentially visual or perceptual in nature. While Anderson's work has been useful for examining components of verbal communication, this study seems to cry out for a theory which explains how learners develop mental prototypes or visual conceptual models as a result of alternative modes of presentation (Dunn, 1983; Tennyson et al., 1982).

One might reasonably ask whether other theories were considered. It is unclear, for example, why the authors chose to explore the relative effectiveness of "varied" and "unified" micrograph content. Was this decision dictated by some unspecified theoretical construct?; by previous research findings?; by classroom experience?

Contributions of the study. Perhaps the strongest contribution of the present study is the support it lends to the growing body of evidence that verbal structure in classroom communication plays an important role in the acquisition and retention of propositional or conceptual knowledge. The results also suggest that kinetic structure is less important in the acquisition of visual knowledge.

Due to possible methodological problems (see Research Design), the finding that micrograph content influences the acquisition of micrograph reading skills seems less certain.

Research design/written report. Careful readers of this paper may find it a frustrating experience. In many respects it appears that the authors took great pains in the design of the experiment and, in general, the study appears to be well-conceived. Unfortunately though, some of the findings are difficult to assess because the written report is uneven in its description of experimental methods and procedures.

For example, it is clear that the kinetic structure variable was well controlled. Readers are treated to a detailed tabular account of the number of discourse units, verbal elements, slides, tape times, and kinetic structure coefficients of the audiovisual presentations. On the other hand, the written report is vague with respect to other, equally-important methodological issues. For example, one might ask:

1. What was the nature of the three practice sessions attended by the students? Were students actually given micrographs to read? How did these sessions differ among the comparison groups?
2. What was the actual content of the micrographs students observed? We are told that the "varied" group observed micrographs of five different protozoan species (Blepharisma, Stentor, Paramecium, Euplotes, and Tetrahymena) and that the teaching objectives of the "varied" and "unified" (Tetrahymena) presentations were judged to be comparable. However, the objectives were never identified, and therefore the reader can only guess about the structural and ultrastructural elements actually presented to the students.
3. Why were the students not pre-tested? Random assignment to treatment groups is acceptable practice as a minimal standard in educational research but pre-testing might have helped identify additional sources of variance in micrograph reading ability. Perhaps differences in prior knowledge of relevant concepts or skills or maybe differences in learning styles (such as field dependence/independence) might have proven as important as kinetic structure and micrograph content.
4. Most importantly, which specific concepts and skills were addressed in the post-test? We are given information about the form of the post-test items (short answer and multiple choice) and the alpha reliabilities of the subscales. However, with respect to content, we are only told that subscale G measures "general concepts" and subscale

RM is "related to reading the given micrograph." What were the questions? How was the instrument scored? What levels of cognitive difficulty were assayed? What structural or ultrastructural features were depicted in the "given micrograph"? How similar was the "given micrograph" to those observed in the treatments?

Validity. The validity of the study is difficult to evaluate because the authors are vague in their description of critical methodological issues. If, for example, the "given micrograph" depicted Tetrahymena or features common to it, then clearly students receiving the "unified" treatment would be expected to have an advantage; an advantage not necessarily derived from the treatment per se but rather from familiarity with the test item stimulus.

In the "Implications" section of the paper, the authors tell us that,

> " ... when students are expected to perform a learning task that requires identification of objects presented in the lesson, the instructional material should be as similar to the test material as possible."

This statement seems to suggest that students in the "unified" treatment group may have been provided with instructional materials that were more similar to the test material than those in the "varied" treatment group. If so, it would seem to call the conclusions of the study into question and to suggest that rote learning might account for observed differences in micrograph reading skills.

Future research. Research on the structure of classroom communication is especially timely now as cognitive psychologists explore the structure and use of knowledge by experts and novices in a variety of domains (Chi et al., 1988). Among the domains that have been investigated are problem solving in physics and genetics, diagnostic skills in medicine, chess playing, typewriting, computer programming, judicial decision making, and memory for restaurant orders.

One of the many findings these workers report is that individuals who possess expertise in a discipline display a "more cohesive and integrated" knowledge structure than beginners (Gobbo and Chi, 1984). Accordingly, we now have evidence that highly cohesive instruction enhances achievement and, in turn, that high levels of achievement reflect a cohesive knowledge base.

With these studies as a starting point, perhaps one area of research that might prove fruitful is the relationship between communication structure and cognitive structure. Using such techniques as "concept mapping" (Novak and

Gowin, 1984) and clinical interviewing, it might be possible to document the effects on cognitive structure of lessons possessing varying levels of kinetic structure. If such a relationship could be established, it would suggest that at least some important features of expertise are teachable by direct methods.

REFERENCES

Ali, A. "The Effectiveness of Laboratory Photomicrography for Studying Microbiology in Nigerian Secondary Schools." Journal of Research in Science Teaching, 21 (1): 63-70, 1984.

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

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

Anderson, O. R. "Research on Structure in Teaching." Journal of Research in Science Teaching, 11: 219-230, 1974.

Chi, M., R. Glaser, and M. Farr (Eds.). The Nature of Expertise. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers, 1988.

Dunn, C. S. "The Influence of Instructional Methods on Concept Learning." Science Education, 67: 647-656, 1983.

Gobbo, C., and M. Chi. "How Knowledge is Structured and Used by Expert and Novice Children." Cognitive Development, 1: 221-237, 1986.

Montgomery, R. J. "The Relative Effectiveness of Three Levels of Realism in Visual Illustrations on Achievement of College Biology Students." Unpublished Doctoral Dissertation, University of Maryland, 1979.

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

Tennyson, R. D., J. N. Chao, and J. Youngers. "Concept Learning Effectiveness Using Prototype and Skill Development Presentation Forms." Journal of Educational Psychology, 74 (6): 329-334, 1982.

Horton, P. B., R. H. Fronk, and R. W. Walton. "The Effect of Writing Assignments on Achievement in College General Chemistry." Journal of Research in Science Teaching, 22 (6): 535-541, 1985.
Descriptors--*Academic Achievement; *Chemistry; *College Science; *Content Area Writing; Higher Education; Science Education; *Writing Composition

Expanded abstract and analysis prepared especially for I.S.E. by F. Gerald Dillashaw, Bradley Universiy.

Purpose

The purpose was to investigate the effect of additional writing assignments on achievement in a college general chemistry course. Specifically, the authors investigated whether writing summaries of lectures would lead to higher achievement on unit exams.

Rationale

This study was initiated because of concern over writing abilities of college students and the premise that the teaching of writing is the responsibility of all teachers regardless of discipline. However, the authors found little empirical support for their study in the literature. Hence, this study is an attempt to contribute to knowledge in this area.

Research Design and Procedure

The researchers used a pretest-posttest control group design. Subjects consisted of 64 students enrolled in the second quarter of a general chemistry course. The subjects were randomly assigned to one of two treatment groups. The experimental group was required to submit summaries of eight consecutive lectures following instructions provided by the instructor. The students were told that the purpose of the assignments was to allow the instructor to spot any misconceptions from the lectures. The students were also told that clarity of writing would be used in the grading of the summaries. The control group was told that they would be required to turn in summaries of the eight lectures following the first eight. The assignments of the control group were collected but not used in the data analysis.

The lecture summaries were begun the fourth week of the quarter to give enrollment time to stabilize in the course. At the end of the fourth week the first exam was given and this exam was used as the pretest. The experimental group then summarized the eight lectures following the first exam. Feedback concerning any misconceptions of the chemistry concepts was given to the entire class. The second exam was given and these scores were used as the dependent measure. The data were analyzed using ANCOVA with the first exam scores (pretest) as the covariate.

Findings

Results of the study indicated that the experimental group scored significantly higher than the control group on the posttest exam. The experimental group scored an average of 7% higher on the posttest when compared to the control group. A short (5-item) survey of the group indicated that 83% of the students felt the technique had increased their understanding of the course material.

Interpretations

The authors suggest that the study is limited in that it was only of three-week duration. Also, since the control and experimental groups were in the same room, both may have benefited from the feedback. Another limitation pointed out by the researchers is the use of only one class and one instructor. Further, the results may be due to the fact that the subjects in the experimental group spent more time on the material to prepare the summaries rather than the act of writing the summaries. They suggest that further study is needed.

ABSTRACTOR'S ANALYSIS

The researchers have attempted to provide additional data that would help to understand the relationship between writing and achievement. The researchers had little support from the science education literature to guide their study;

thus, they were essentially starting from the beginning to study the relationship between writing and achievement in science. However, it seems that the researchers did not seek much information about the relationship between writing and achievement from other disciplines to determine if there was some evidence to support the notion that such a relationship might exist in any area.

The researchers had stated that there is general and growing concern over college students' writing abilities and that all instructors have the responsibility of teaching writing. The abstractor agrees with the researchers, but this study addresses this concern only in a narrow way. The use of written summaries of lectures is a fairly limited type of writing assignment. Although such summaries may assist students in understanding concepts presented in lectures, such assignments may have limited value in teaching students skills in expository or critical writing where the student must generate more original material that is called for in a summary. In addition, the researchers failed to provide information regarding the improvement, if any, in the writing skills of the students.

For an initial investigation into this area, the research design and procedures were adequate. The researchers paid careful attention to the requirements for an experimental study and addressed most of the concerns that would potentially confound the results (Gay, 1987). As the researchers pointed out, the limitations in a study of relatively short duration and a single instructor may be potential explanations for the results.

The researchers are correct in suggesting that the simple fact that the students had to spend more time on the material to write the summary may also be an explanation for the outcome rather than the act of writing itself. In fact, this possibility may be the most significant. It is not clear how long or how detailed the summary had to be. Depending on the exact requirements, the students may have had to spend considerable time with the content of the lectures, hence were more academically engaged with the material as compared to the control group. This additional time could easily result in increased understanding of the material independent of the process of organizing the material for the writing assignment. This aspect of the research is one that needs additional study.

It is clear that the relationship between writing and achievement is a complex one and that further study is needed in view of the concern about students' writing abilities.

REFERENCE

Gay, L. R. Educational Research, 3rd Ed., Columbus, Ohio: Merrill Publishing Company, 1987.

Isom, F. S. and R. E. Rowsey. "The Effect of a new Prelaboratory Procedure on Students' Achievement in Chemistry." Journal of Research in Science Teaching, 23 (3): 231-235, 1986.
Descriptors--*Academic Achievement; *Chemistry; *College Science; Higher Education; *Science Instruction; *Teaching Methods

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

Purpose

This study compared a small-group, interactive method of preparation for freshmen chemistry laboratory with the traditional prelaboratory lecture. It tested the hypothesis that students assigned to the experimental group, who participated in the alternative method, called Prelaboratory Preparation Period (PLPP), would have higher scores on post-laboratory measures than those assigned to the control (lecture) group.

Rationale

PLPP is a method of laboratory preparation, developed by the authors, that incorporates small-group instruction followed by opportunities for students to ask questions of the instructor and to discuss the experiment to be done among themselves. This process, which takes place a day or two prior to the laboratory work, is scheduled for approximately 25 minutes of instruction and 20 minutes for interaction. Advantages cited are the opportunity for informal interaction between students and instructor and among students themselves and the provision of time between the PLPP and the laboratory which allows students time to think about questions or problems related to the work to be done in the laboratory. This is, in effect, an advance organizer for the laboratory.

Research Design and Procedure

A posttest only, control group design was used to test the hypothesis referred to above. Freshman laboratory sections were randomly selected over four quarters and designated experimental or control. The total number of sections was eight: five experimental and three control with a total sample of

233 subjects. Experimental classes received instruction via PLPP prior to performing the laboratory exercises; control classes met in a group of 48 and had a 20-minute preparatory lecture immediately preceeding the laboratory. Data were collected from all groups for seven laboratory exercises as follows: Determination of Density, Avogadro's Number, Gravimetric Determination of Water of Hydration, Composition of a Compound, Determination of Gas Constant, Determination of Molecular Weight, and Stoichiometry of Acid-Metal Reaction.

The data consisted of laboratory reports and scores on short quizzes. Reliabilities (K-R 20) of 0.78 to 0.84 are reported without specifying exactly what kinds of data were included in the reliability calculations.

Two statistical procedures were used to test the null hypothesis of no difference between experimental and control groups: a Multivariate Analysis of Variance producing a Wilks Lambda Omnibus F-test followed by a series of Univariate F-tests.

Findings

The Omnibus F-value of 2.40 exceeded the critical value for a probability level of 0.05, leading the authors to reject the null hypothesis of no difference between the experimental and control groups. The authors offer this as proof of the superiority of the PLPP. Further analysis showed that the presumed positive effect of the PLPP was dependent on the particular laboratory exercise under consideration since the Univariate F-test showed that there was a significant difference ($p = .05$) between experimental and control groups in only one of the laboratory exercises.

Interpretations

The authors assert that their findings support two conclusions, namely, that the PLPP significantly improves academic performance of general chemistry students at the university in question and that the effect of PLPP varies according to the nature of the laboratory experiment. They believe that PLPP is a useful alternative to the traditional prelaboratory lecture.

ABSTRACTOR'S ANALYSIS

At a time when the laboratory is in danger of becoming extinct in high school and is under increasing time pressure at the undergraduate level, one cannot but welcome a study that gives serious consideration to improving the quality of laboratory work and, thereby, increasing student achievement. We can only applaud the effort to provide students with an advance organizer that will help them to understand what they are to do, give them opportunities to ask questions and give them time to think about what they are to do. All of this is in contrast to the prelaboratory lecture with which we are familiar, both as students and instructors. The lecture is often rushed, there is seldom time for meaningful questions, and the explanations come before the students know what needs to be explained. An alternative to this outdated method is surely needed.

Although we would like to believe that the authors' alternative is superior to the traditional prelaboratory lecture, there are several problems in the design and in the description of the study that prevent the results from being as clearcut as one might wish.

The posttest-only design depends heavily on the random assignment of subjects to groups, a condition that was not met in this study. This lapse would not have been as serious if we knew how students were assigned to laboratory sections, but it is unlikely that the assignment was random since sections are often formed on other bases. One common biasing factor is students' schedules; for example, all math majors may have to take a course that conflicts with one laboratory section with the result that they are assigned to other sections. True randomization is often not possible in an experiment of this kind but its lack is a threat to validity, nevertheless.

No explanation is given for the fact that there were five experimental sections and three control sections, making an imbalance in the numbers of control and experimental subjects. It would have been preferable, though perhaps not possible, to have had an equal number of sections assigned to each treatment. In that case, it might have been preferable to use the sections rather than individuals as the unit of analysis.

Another problem that was not taken into account was the difference in the amount of time spent in the experimental and control prelaboratory preparation periods. The PLPP took place over a 45-minute period while only 20 minutes were allowed for the lecture. It could be argued that the difference in time was a significant factor.

A difficulty for the reviewer is that the explanation of the data sources does not tel us how the final scores used in the analyses were formulated. We are told that laboratory reports and grades on quizzes were used, but we are left in the dark as to how these were combined or weighted; thus the origin and meaning of the scores is unclear. Since the purpose of the study was to measure the effect of PLPP on academic achievement and the scores in question were the dependent variable, we must assume that this score is the measure of "academic achievement." A more precise definition is needed here.

At first glance it seems surprising that the overall F-test allows us to reject the null hypothesis while the individual, post hoc, analyses indicate only one laboratory exercise for which the null hypothesis can be rejected. This is somewhat less surprising when the data are examined closely, for there are two cases, one with a very large MS, that are close to the cutoff point for significance. The authors offer an explanation for the large differences in effect of PLPP on the several laboratory exercises but the proffered explanation is somewhat disingenious and not very convincing. The lack of difference between experimental and control groups in all but one of the laboratory exercises is, in fact, rather damaging to the argument that the experimental treatment is superior to the traditional approach.

Summary. The effect of laboratory experience on student learning and the search for ways to increase the positive outcomes of laboratory work are areas of considerable interest and areas in which new knowledge is much needed. The authors of this report have a good idea that has much intuitive appeal and is certainly worth pursuing. The present study is, however, more in the nature of a pilot study than of a finished piece of research. The authors have developed a promising procedure that probably could be used without substantive change in a future study, either by themselves or other researchers. If this were to happen, very careful attention should be paid to random assignment of subjects to groups or to blocking of subjects to produce representative groups; to definition of the dependent variable and a detailed description of the method assigning scores; and to the choice of statistical procedures.

It is very difficult to design and execute a study in the real world that stands up to close scrutiny but, without attention to important details of design and analysis, the results remain inconclusive.

Stanley, Julian C. and Barbara S. K. Stanley. "High-School Biology, Chemistry, or Physics Learned Well in Three Weeks." Journal of Research in Science Teaching, 23 (3): 237-250, 1986.

Descriptors--*Academic Achievement; *Biology; *Chemistry; *Course Descriptions; *Physics; Science Education; Secondary Education; *Secondary School Science

Extended abstract and analysis prepared especially for I.S.E. by Constance M. Perry, University of Maine.

Purpose

The purpose of the study was to determine if intellectually highly able youth can successfully learn the equivalent of one year of high school biology, chemistry or physics in three intensive weeks of study.

Rationale

There is much concern about education in America as evidenced in literature (Boyer, 1983; Goodlad, 1984; Sizer, 1984). Part of that concern relates to the need for improvement in the teaching/learning of science in our schools. Much of such improvement efforts have addressed better preparation and pay for teachers; however innovation in the organization of science courses is another potential avenue for improvement. Scientifically highly talented girls and boys need special opportunities to advance faster in the basic high school science courses than regular high schools make possible, in order to keep them from becoming bored with science and to provide them with more and earlier science opportunities. The end result may well be an increased number of doctorates in science, and engineering, of which there is an undersupply (Begley, 1985).

Since June, 1972, the Study of Mathematically Precocious Youth (SMPY) at the John Hopkins University has examined a variety of ways to teach high school math faster and better to highly able students. Their attempts have met with excellent results (Bartkovich and Mezynski, 1981; Benbow and Stanley, 1983; Stanley and Benbow, 1986). A similar study looking at science and highly able students is reported here.

Research Design and Procedure

In the summer of 1982, 25 highly capable youth, ages 11-15 studied biology and 13, ages 12-15, studied chemistry intensively for 3 weeks. All had been required to score at least 500 on the mathematical part of the College Board's Scholastic Aptitude Test (SAT) before age thirteen, and to score at least 930 on the combined math and verbal parts.

Biology was taught from a standard textbook by a retired high school teacher who had taught Advanced Placement (AP) biology for many years. Chemistry was taught by a high school teacher without AP experience. The classes met five hours a day for 15 days and included some laboratory.

The biology students were pre and posttested using different forms of the College Board biology achievement test. The chemistry students were posttested only, using the appropriate College Board test, because chemistry has a particularly specialized vocabulary and basic concepts making pretesting inappropriate. A questionnaire was administered to determine student satisfaction with the course, at its completion.

In the summers of 1983-85 the program was expanded to include phsyics. Pre and posttests were administered in each subject. As in 1982, all students were required not to have had a prior formal course in the science subject.

Mentoring by mail was offered during the academic year after the 1982 summer program to students who did not have access to a college-level advanced placement course. Ten of the 25 biology students and two of the 13 chemistry students enrolled.

Correlations were computed amongst six biology class variables: pretest score, posttest score, AP exam score, age in months, mentored or not, and sex. An AP score of 2 was arbitrarily assigned to the eleven who did not take the AP biology exam, since without a score of ≥ 3, college credit would not be granted. Correlations were not computed for the 1982 chemistry group because of the small number of students (13). Correlations are not mentioned for the 1983-85 groups.

Findings

1982 biology class. The pretest scores ranged from 420 (14th percentile) to 690 (89th percentile) with a mean of 565 (54th percentile), median of 560

(52nd percentile) and a standard deviation of 71 (vs. 109 for the norm group). Item analysis of the pretest responses showed that the students tended to answer correctly items requiring reasoning ability and to incorrectly answer or omit items requiring specific technical knowledge.

The posttest scores ranged from 590 (61st percentile of national norms) to two 800s (30 points above the minimum 99th percentile). The median was 727 (95th percentile), the mean 721 (94th percentile), and the standard deviation 57. The mean score in the 18 days from pretest to posttest rose 156 points. Because most of the students had already taken the biology test prior to the pretest, little of the gain from pretest to posttest is due to further practice effect. (The mean gain from pre-pretest was itself only 20 points).

Ten of the 25 students enrolled for the assistance of mentoring by mail in order to prepare for the AP exam. The students varied in their efforts from doing all the work to not opening any of the mail. Nine of the ten mentored students and five of the fifteen not-mentored took the AP exam. The mentored scored 13, 34's and 55's and the not-mentored scored 23's and 35's.

Correlations amongst six biology-class variables listed earlier yielded significant results. Only being-mentored made a statistically significant contribution (r=0.54) to estimating AP score. Two other correlations were significant. They were 0.48 between pretest and posttest scores and 0.42 between age and posttest score.

<u>1982 chemistry class</u>. The posttest scores of the 13 students had a median of 743 (95th percentile) and a mean of 734 (93rd percentile) with two 800's. The range or standard deviation was not reported. Only two students took the AP exam in 1983 or 1984. One scored ·, the other (1 of the 2 mentioned) scored 3; yet 70% of the biology class took the equivalent biology test and averaged 4.42.

<u>Classes two years later</u>. The pretest and posttest median scores only are reported for classes during the summer of 1984.

Table I

Biology, Chemistry, and Physics Classes, Summer 1984

	Classes 1st 3 weeks			Classes 2nd 3 weeks		
	N	Pretest Median	Posttest Median	N	Pretest Median	Posttest Median
Calculus-prerequisite physics	___	___	___	23	537	680
Non-calculus prerequisite physics	23	547	650	___	___	___
Chemistry	24	465	625	24	450	642
Biology (2 sections each term)	47	467	591	4[illegible]	432	629

Interpretations

In analyzing the questionnaire it was noted that, in 1982 courses, the biology teacher had much more personal appeal, was able to engender more enthusiasm for the subject, and had more knowledge of the subject at the college level. From this information the authors state that it may be inferred that the teacher differences were responsible for the sizeable AP difference between biology and chemistry in percentage of students taking the exam and the mean scores.

The authors state that:

> Clearly, this method for moving scientifically apt students through the basic high school course in biology, chemistry, or physics fast and well saves many of those youths, their schools, and their parents' time, patience, and money. It represents a novel supplementation of secondary education by a university ... Even more important, however, is the student's zest for the subject that the accelerated course can provide. (p. 243)

The authors suggest that the three weeks is not sacred. Perhaps six weeks would be better. Saturday mornings might better be used academically rather than socially as was done in the three week residential course.

Proper handling of laboratory work was a problem. The 1982 students complained of not enough lab time. In 1985 a lab was scheduled every afternoon which cut deeply into the instructional time. Authors state that more experimentation with lab types and lengths should prove helpful.

In conclusion, Stanley and Stanley assert that:

> To achieve desirable progress in a society steeped in technological and scientific pursuits, science must attract the best minds....A faster pace and more stimulating approach to teaching the ablest are possible. They may also be necessary lest keen-minded young people lose interest in areas where they are particularly capable of making significant contributions. (p. 245)

ABSTRACTOR'S ANALYSIS

The results of the study - extremely high College Board achievement test scores - definitely show that highly capable science students can very successfully complete the equivalent of a year-long high school science course in three intensive weeks. In addition, many were able to go on to earn college credit through advanced placement exams by taking a subsequent course in their own school or by the mentoring-by-mail program.

The results are very important, but perhaps as important as the fact that highly able science students can learn well in three weeks what is traditionally offered in a school year, is the realization that if science can be successfully taught that way, so perhaps can other courses. Also, the study raises the question: If three week courses work, what other structures might work other than year-long course designs, and for whom would they be effective? Students, parents, school systems, and universities are encouraged to look beyond the traditional structure to other options for the delivery of high school courses to the intellectually highly able student. A not so subtle message is sent in the article, when the authors quote Stanley and Durden (1983) in regard to especially apt students in junior and senior high schools..."students who are age-in-grade studying each school subject 45 or 50 minutes daily with regular, pedagogically certified teachers operating in the usual way seems unnecessarily limited and perhaps futile...." (p.6&8)

The study was interesting to read and based on a strong rationale for improving the education in science, especially for the intellectually highly able one-half of one percent of the junior/senior high school population. However, the report was somewhat choppy and therefore confusing. To find the criteria for acceptance of students into a 3-week course one was required to read several different sections, and to find out if any 1982 chemistry students were mentored-by-mail, one had to look in parentheses under a heading other than "Mentoring by Mail."

The study is an important and logical extension of research concerning intellectually capable math students succeeding in shortened-in-time versions of math classes. The results strengthen the idea that intellectually highly able students benefit from a change in structure of courses in that they can advance more quickly while still achieving very well on College Board achievement tests.

Although the study answered the question whether highly able students can complete successfully a traditional year-long science course in three weeks, it left several other questions, perhaps of somewhat lesser importance but questions albeit, unanswered.

The reader does not know how these same or comparable students would do on the same tests after taking a traditional 5-day a week, year-long course. Such a comparison could lend strength to the study if the 3-week scores were higher or equal to the year-long scores. Since the student scores after the 3-week courses were so high, such a comparison might not yield any valuable information but one does not know that. Further study in this area could prove interesting.

The authors stress that, for the intellectually highly capable science student, a year-long traditional science course would be very boring and that the 3 week course enables the students to learn science with more zest. The three week course is labeled as more stimulating. The statements are reasonable and very likely true but the reader is provided no data to back up the statements. One must assume that the data have come from the past experience of the authors with talented youth and/or the questionnaire the students completed. More specific reference to the data source and a copy of or a list of questions from the questionnaire would be helpful to the reader. In addition, a clarification as to why the three week course is deemed more stimulating would be beneficial. Is it because it is shorter, more intensive, and therefore less boring, or is it the teacher that increases the amount of stimulation the course contains, or is it a combination of the two?

Some of the other questions raised by the reporting of the study are listed below. Is science always boring for the intellectually highly capable student if taken in a traditional year-long program? That appears to be the assumption made. Aren't there high shool science teachers who can work with such students and not bore them? Is the assumption fair?

In the article, a section titled "teaching methods" described the characteristics of the most effective teacher for the 3-week science program. Where did that information come from, past experience, the questionnaire, or what?

It appears that in the 1983-1985 program, chemistry students were given a pretest, yet in 1982 the article states that a chemistry pretest was deemed inappropriate. What happened? Is the pretest now given regularly, and - if so - why?

Why weren't the data for the 1984 group as complete as for the 1982 group?

The authors deserve credit for raising several questions beyond the scope of the reported study. A few examples are listed.

When the lab time was dramatically increased in 1985, what were the results on the posttest scores and how did the increased lab time affect questionnaire responses?

Were there marked differences in the math and verbal scores on the required SAT? Students had to have scored a 500 on math and a combination of 930 on math and verbal combined to qualify for the program. Were there students with 600's in math and 350's in verbal? If so, did those score differences affect achievement test scores on pre and/or posttests?

Are intellectually highly able students who move through science coursework more rapidly than is traditional more likely to pursue higher degrees and careers in science than their counterparts who follow a traditional timetable?

The study is exciting, both in what it answers and what it doesn't. The findings deserve attention by policy makers who have the power and ability to effect change in the education of our students, especially our highly capable ones.

REFERENCES

Bartkovich, K. G. and K. Mezynski. "Fast-paced Precalculus Mathematics for Talented Junior High Students: Two Recent SMPY Programs." Gifted Child Quarterly 25 (2): 73-80, 1981.

Begley, S. "Wanted: More Graduate Students in Science and Engineering." Johns Hopkins 36 (4): ii-ix, 1985.

Benbow, C. P. and J. C. Stanley. Academic Precocity: Aspects of its Development. Baltimore, MD: John Hopkins University Press, 1983.

Boyer, E. L. High School: A Report on Secondary Education in America. New York: Harper and Row, 1983.

Goodlad, J. A Place Called School: Prospects for the Future. New York: McGraw Hill, 1984.

Sizer, T. Horace's Compromise: The Dilemma of the American High School. Boston: Houghton-Mifflin, 1984.

Stanley, J. C. and W. G. Durden. "Supplemental Teachers of Science and Mathematics." Science 220 (4592, April 1): 6 & 8, 1983. Letter to the editor.

Tobin, Kenneth. "Effects of Extended Wait-time on Discourse Characteristics and Achievement in Middle School Grades." Journal of Research in Science Teaching, 21 (8): 779-791, 1984.

Descriptors--Abstract Reasoning; *Cognitive Processes; *Elementary School Science; Intermediate Grades; Junior High Schools; *Questioning Techniques; Science Education; *Secondary School Science; *Science Instruction; Teacher Student Relationship; *Verbal Communication

Expanded abstract and analysis prepared especially for I.S.E. by Linda R. DeTure, Rollins College.

Purpose

The purpose of this study was to investigate the impact of extended teacher wait time in whole class settings on a number of teacher and student outcome variables. The author asked four research questions regarding: the variation in the length of wait time of teachers in the experimental and control groups; the relation of teacher wait time to both teacher and student discourse variables; the relation of teacher wait time to student achievement; and the relation of the student and teacher discourse variables to student achievement.

Rationale

Tobin set the context of the rationale in the paradigm of learning, rather than instruction, using information processing theory as the model. Upon noting the elements required for learning higher level thinking skills utilizing this model, he focused on how the teacher can behave to enhance the student's cognitive processing. Because students need time to manipulate and process information being stored and/or retrieved, the teacher should incorporate pauses or silence at appropriate times during discourse to allow for the processing. Providing the student with time to think during the lesson should result in several changes including improved achievement as measured by objective tests.

Based on Rowe's and others' research of the dynamics of wait time, Tobin expected to refine and add to the general body of knowledge concerning the role of pausing in instruction. In particular he examined the instructional level encountered the most in teaching, i.e. whole class or large group. Many previous studies were conducted in small groups or microteaching settings.

A premise of the study is that in whole class settings a major goal of silence is to encourage thinking as opposed to student-student interactions, a common goal of small group instruction. Tobin focuses on a species of wait-time designated as "teacher wait time" which is solely under the control of the teacher rather than "student wait time" which is controlled by the teacher and student. It should be noted these do not correspond to Rowe's definitions of wait time I and II. Since Tobin was seeking to validate strategies to enhance thinking at higher cognitive levels, he selected probablistic reasoning as the topic of the seven lesson sequence.

Research Design and Procedure

The sample for the study was twenty intact classrooms of sixth and seventh graders. Ten of the classes were randomly assigned to the treatment group whose teachers received feedback about their wait time as a means of training. The remaining ten classes received no wait time feedback. The treatment continued over a series of seven lessons provided by the researcher to teach the concept of probability.

The discourse variables measured for both teacher and student were: structuring; soliciting; responding; and reacting. The teacher's reacting moves were reanalyzed according to four criteria; low level, high level, mimicry and probing (a form of soliciting). The length and frequency of student and teacher utterances were also measured.

A test of formal reasoning was used as a pretest, and a ten-item objective achievement test was used as the post test. Data for wait time and the student and teacher discourse variables were collected from audiotapes recorded during each of the lessons. Five minute segments extracted near the beginning and end of the lessons were used as the data sources for the variables.

The basic overall design for the discourse variables is a posttest only, control group design. However, when analyzing the achievement gains, the Test of Logical Thinking (TOLT) is used as a pretest and a covariate for the gain scores making is a pretest-post control group design for those variables. For the wait time and the discourse variables, the means and standard deviations averaged from the seven lessons were used as the units to compare the two groups with an analysis of variance, and an analysis of covariance. The relation between the student variables and summative achievement was determined by multiple regression analysis.

Findings

A significant difference in the average wait time was found between the two treatment groups. Over the seven lessons sequence, the wait time of the treatment groups increased in a lower trend from 1.9 seconds to 4.4 seconds with an average of 3.2 seconds. No such trend was noted for control group whose wait time averaged 0.9 seconds.

Of the eight discourse variables, six showed a significant difference between treatment groups. For the wait time feedback group the average length of student discourse was greater (4.7 seconds to 2.6 seconds) and the average length of teacher talk was less (6.0 seconds to 8.4 seconds). Corresponding to these changes are the number of utterances which decreased from 55 to 35 per 5 minutes periods for the wait time group and remained at about 65 for the non-wait time group. While the proportion of teacher solicitations was greater for the treatment group, the number of questions decreased from an average of 4 per minute the first lesson to 2 per minute the seventh lesson. The other group remained constant at 4 per minute. The proportion of teacher reacting was less and the proportion of teacher structuring was greater for the wait time group.

When breaking down the reacting moves, it was found that probing was significantly higher for the wait time group and that mimicry was higher for the control group. As a trend for the wait time group, mimicry decreased from 38 to 13 percent and soliciting increased from 26 to 59 over the seven lessons. No differences were found for either low or high level reacting. The analysis of covariance indicataed that the summative achievement for the treatment group was significantly higher than for the no treatment group. Thus, teacher wait time was significantly related to 9 teacher and student discourse variables.

Multiple regression showed that formal reasoning, length of student discourse, and the proportion of student reacting moves were all positively related to summative achievement.

Interpretation

Tobin found wait time to be easily manipulated for most teachers as long as feedback was forthcoming. Having extended wait time relating significantly to nine outcome variables supports the thesis that increased silence is being used by both teacher and student for cognitive processing. The results indicate that the use of extended wait time in middle school grades can improve student achievement when holding variations in formal reasoning ability constant. This research extends the findings of improved achievement to whole class settings.

The only student discourse variable significantly related to the treatment group was length of discourse. Apparently interrupting students less enabled them to process information in a more effective manner, leading to the suggestion that teachers adopt strategies, such as silence, to increase the average length of student response. Most of the changes resulting from increasing wait time were in teacher behavior. For example, when teachers probed more, the students had to react to and evaluate the responses of others. In classes where students reacted more, achievement was greater. The overall implication is that teachers could utilize wait time in combination with redirecting and probing strategies to increase the length of student responses and to increase the amount of student reacting. Achievement as measured by tests would then be enhanced. Tobin notes that extended wait time does not in and of itself lead to improve achievement. He asserts that the corresponding changes in teacher and student dialogue lead to higher achievement and that teacher training should focus on a wide range of teacher behaviors including pausing. Wait time alone may be necessary, but not sufficient to bring about the desired changes.

ABSTRACTOR'S ANALYSIS

This study is an excellent example of what could be called a second generation wait time study. If Rowe's pioneering research, which identified the phenomenon of pausing, is considered the first, this and simil. studies might be considered the second. In the first set of wait time research studies, Rowe identified pausing as being a variable that contributed to higher levels of inquiry. Based on this discovery, she and other conducted numerous studies to find some of impact using extended silence. By and large her findings have held up to the scrutiny of the second generation studies, particularly regarding the time of silence needed for changes to occur (approximately 3 seconds) and the places it was most important for teachers to pause (after a student's response). Many of the early studies were conducted in small groups or microteaching settings with elementary science students.

The second generation of studies keyed into exploring the conditions under which using wait time would be effective. Topics extended to levels of instruction, various subject areas such as math, student achievement as an outcome, and teacher training. Tobin's work has contributed a major portion of that body of knowledge and has helped point the way to a new generation of studies about learning in the classroom, where wait time is but one of a myriad of variables needed to bring about effective change in the classroom.

A few points need to be made in regard to this particular study. Generally the study was well executed having many of the features of a "model" research plan. Togin provided a rationale for the study grounded in learning theory and supported by the evidence of a number of research studies. The variables were operationally defined with the research plan being a traditional Stanley and Campbell design for an experimental study. The analyses of variance, covariance and multiple regression were appropriate methods for analyzing the data.

Even so there are a few deletions and questions that need to be addressed or clarified. Close examination of the discourse variable shows some overlap. For instance, the distinction between coding the probing and soliciting is not clear. It appears that when a teacher reacted to a student response with a question it was coded as soliciting. It is unlikely that the questions were coded as both a reaction and a solicitation. If it were coded as a reaction, then soliciting would be undercoded and reactions would be overcoded and vice versa. Either way it would impact measuring the two variables and should have been discussed. In the results section the proportion of soliciting was found to be higher and the proportion of reacting was lower for wait time feedback groups which appears to indicate that probing was coded as soliciting. A sentence or two as to how coding was done could have cleared up the question.

In terms of measuring the wait time and the discourse variables, Tobin indicated that two five-minute samples, one near the beginning and one near the end of the lesson, were selected for analysis. Initiating and closure discussions are often different in nature from mid-lesson dialogue and would not be representative of whole class dialogue. Two or three random samples taken throughout the one hour lesson may have yielded somewhat different results and would have been more valid as representative of whole class instruction.

The author measured only 30 randomly selected wait times for 10 minutes of recording instead of averaging all the wait times which is commonly done by wait time research. Given the nature of the servo chart recordings, it would be interesting to know how the 30 were randomly chosen. The question is one of methodology rather than concerning the validity of using random measures. The researcher noted that teacher utterances, also a sample of 30, were randomly selected from the beginning, middle and end of the lesson. Perhaps he did use three samples for wait time and the omission referred above is one of reporting i.e., he did not take a mid lesson sample, rather than of procedure, but it is not clear as to which occurred.

Two important sets of descriptive data were missing from the report. No standard deviations for any wait time means were listed in the report or tables. An indication of the variance between the groups is needed to determine how much the two groups differed from one another. Discussion is warranted because it deals directly with the first research question which provides the framework of the study. An omission that may be more important is the absence of reported means and standard deviations for either the summative achievement test or the TOLT. Most readers of a study like this would want that information in addition to the summary table for the analysis of variance, multiple regression, etc. Again the data are directly related to the third and fourth research questions.

When comparing the relationship between wait time and discourse variables, it is likely that some of the changes in the discourse variable are more of a function of their relation to one another than to wait time per se. For example, the number of teacher or student utterances is directly related to the length of utterances. When the teacher uses silence and does not interrupt the students' dialogue then, of course, the length of the utterances will increase and in a finite time period the number of such utterance will decrease. In essence it is two ways of looking at the same data. For research purposes, the author may have a use for the two sets of numbers, but the variables are not distinctly different variables.

Many of the preceding concerns point toward methodological discrepancies when in fact this may be more of a reporting issue. Overall the paper is written in straight forward readable prose but an omission or two in an attempt at brevity can lead to unanswered questions pertaining to both methodology and research design.

One research concern that deserves some attention is how researchers' conceptualize their variables and in this case how the author conceptualizes wait time. Readers of studies about wait time should pay attention to how wait time is being defined and measured because several different approaches are used in the research. With the following dialogue pattern as an example "1 T 2 T 3 S 4 T 5 S 6 S 7 T 8", using either Rowe's or Tobin's definition would result in fairly different measurements of the variable. If the numbers represent an opportunity for silence, Rowe would define pauses 2, 3, 5 and 8 as wait time I and pauses 4, 6, 7 as wait time II. Tobin, on the other hand, defines teacher wait time as the pause preceding any teacher utterance and either the teacher or student could precede. Therefore, Tobin would measure pauses 1, 2, 4 and 7 as teacher wait time and does not account for pauses 3, 5, and 6. Rowe, Tobin, Swift and DeTure have all argued that the pause after a student's response is the most important for bringing about the desired changes in research variables; but each conceptualizes and measures slightly different pauses. Regardless of whether this conflict is resolved or whether it needs to be, it is important for researcher to operationally define wait time in order for the reader to know what is being measured and to make independent interpretations. The issue is not just a question of semantics, partly because not all pausing time is controlled by the teacher and the amount of control may vary depending on the level of instruction. In whole class instruction, as in this study, the teacher may spend more time structuring and controlling the discourse than in small group instruction where a good bit of student to student dialogue is encouraged. To be able to generalize the results, the reader needs to know what was measured.

All in all, this study was well executed and made several important contributions to the body of the wait time research at the time it was conducted. Looking at the role of pausing in whole class instruction was extremely significant. Identifying the use of wait time as a means of maintaining a class environment to enhance learning with middle school children is important. The study also provides additional evidence that both teachers and students use the silence to cognitively adjust, modify, and process the information they are sharing during discussions and the end result being improved achievement by the students.

One of the most significant findings of the study may be the demonstration of a linear increase in wait time over the series of lessons. Tobin's teacher training consisted of alerting the teachers to the behaviors and giving specific feedback concerning their pausing after each lesson. Although he did not specify the form of the feedback, i.e., from the actual audiotapes or just observations; the t .ning was effective over time. Training teachers to engage in criterion wait time has been a concern of a number of wait time studies. It would be interesting to see what happened to these teachers' extended wait time in the absence of feedback.

Generally the study contributed to a set of studies, all of which support the importance of using wait time to improve class interaction. The positive effects of using wait time are not nearly so much in question as is how to have teachers engage in wait time at a nonintervention level such as when the feedback and support stops. To use wait time routinely in classroom when the elements of training and support are not present is still a goal to be sought. How to achieve such behavior in the absence of intervention could be one focus for the next generation of studies.

Tobin begins to indicate the way for the future wait time studies with this prophetic statement, "Classroom interactions are complex and differences between teachers and students probably preclude any strategy being effective in all situations." Wait time studies generally support the thesis that incorporating pausing appropriately into classroom interaction is effective for changing behavior and enhancing achievement. But it is only one of a number of related variables that should be researched, and possibly we should be looking at a broader picture. Currently classroom researchers are moving toward studying the whole class setting and school environments to try to understand why teachers do what they do. An impetus for this is that, in the absence of intervention, teachers often revert back to their pretreatment modes of operating. Maybe researchers should focus more on the constraints of a school setting, the concerns the teacher has about teaching, and the beliefs she/he holds as a key to understanding and ultimately changing behaviors in the classroom. It is unlikely that the researchers for the new generation of classroom interaction studies will be satisfied with manipulating a single variable, even an important one like wait time.

Harty, H., K. Samuel, and D. Beall. "Exploring Relationships Among four Science Teaching Learning Affective Attributes of Sixth Grade Students." Journal of Research in Science Teaching, 23 (1): 51-60, 1986.

Descriptors--*Computer Software; High Schools; *Learning Activities; Mathematics Education; *Mathematics Instruction; *Statistical Analysis; *Statistics; *Secondary School Mathematics

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

Purpose

The purpose of this study was to ascertain if four commonly described affective variables:

attitude toward science
interest in science
curiosity about science, and
self-concept of science achievement

were independent variables or interrelated constructs of the same variable.

Rationale

Analysis of the literature provided a number of correlational studies that indicated evidence of some association between pairs of the four affective variables. The open question is: To what extent are these variables related?

Research Design and Procedure

Using a post-test only design, 228 sixth grade students participated in taking four tests:

"Children's Attitude Toward Science"
"Children's Interest in Science"
"Children's Science Curiosity Scale"
"Self-concept of Ability Scale in Science."

The results of the tests were analyzed by comparing gender differences and correlations to ascertain independence of the four variables.

Interpretation

When the test results of boys were compared with the girls of the sample, there was no evidence of differences except in their interest in science. Boys were clearly more interested in science than were girls. The evidence of high correlations between the attitude, interest and curiosity scores indicated that these may indeed be different descriptors of the same variable. The analysis seemed to provide evidence that the self-concept variable is different from that of the other three affective variables.

ABSTRACTOR'S ANALYSIS

In this study, the researchers have provided a helpful addition to the general area of understanding the affective domain and how it relates to student success in science. While the evidence of correlations between pairs of the four variables--attitude, interest, curiosity and self concept--exists in the literature, they have looked at these four variables in the same sample. The design of the study illustrates how a variety of statistical techniques can be creatively used to direct attention to a single conclusion.

Needed now is a conceptual framework to describe rationale for the attitude/interest/curiosity variable and the self-concept variable--what they are, how they function and why they relate to success in science learning. The authors hint that a possible linkage of the attitude/interest/curiosity variable may be in the language or meaning of terms held by students. It is indeed possible that interviews with students would reveal that they perceive questions about attitudes, interests or curiosity as being all part of the same factor.

By noting the issue of meaning of the terms in the tests, the authors have uncovered the urgent need for precise operational definitions of variables in the study of the affective domain. Operationally defined variables woven into a clear conceptual framework will then enable researchers to proceed to study how these variables can be used to enhance student success in science.

Atwood, Ronald K., Anna A. Neal and Ben Oldham. "Developing Classroom Evaluation Materials for SCIS." Science Education, 68: 163-168, 1984.
Descriptors--*Elementary School Science; Elementary Education; Science Education; *Science Tests; *Science Course Improvement Projects; *Test Construction

Expanded abstract and analysis prepared especially for I.S.E. by Barry J. Fraser, Curtin University.

Purpose

This article describes the development of criterion-referenced tests to assess elementary school students' achievement of the aims of the 12 units of the Science Curriculum Improvement Study (SCIS).

Rationale

Because the science curriculum development efforts of the 1960s produced relatively few evaluation instruments for classroom teachers to use, this study attempted to fill this gap for SCIS.

Research Design and Procedure

For each unit, a test writing team was set up including the school system's science coordinator, a university science educator, and two science teachers, all with experience with SCIS. Drafts of items were critiqued by the science coordinator and university science educator in terms of their clarity, content validity, and consistency with SCIS objectives; later, items were checked for readability by a reading specialist. Next, items were field tested with over 150 students in six classes, and item analysis procedures based on the Rasch model were conducted for a random sample of 100 of these students. Finally, following further revisions based upon these item analyses and comments from teachers, a new version was field tested with a second sample of unspecified size as a basis for forming the final version.

Findings

Kuder-Richardon 20 (KR-20) reliability estimates ranged from 0.742 to 0.760 for tests varying in length from 21 to 63 items.

Interpretations

The authors concluded that the tests provide relatively accurate estimates of students' achievement and that the methods of test development (including the involvement of experienced teachers, having draft items vetted, and basing analyses on the Rasch latent trait model) seem to be effective.

ABSTRACTOR'S ANALYSIS

This brief paper makes a worthwhile contribution to science teaching by providing teachers with access to some previously unavailable achievement tests for use with elementary school students studying any of the 12 units of SCIS.

The test development procedures used were comprehensive and carefully carried out, and they appear sound. As far as it is possible to infer from this article, the tests seem to be of good quality and generally useful, although so far their development, validation, and use has been restricted to a relatively small sample from one country only.

Only two minor points of criticism emerge. First, the authors do not discuss the point that the KR-20 reliability coefficient is more suited to use with norm-referenced tests than with criterion-tests. Second, the important issue of economy is not addressed. In particular, why is it necessary to have some tests so long (up to 63 items) in order to provide teachers with an assessment of student achievement of one content area? Also, for a criterion-referenced testing system, wouldn't it be more informative to divide long tests into a number of reliable subtests which could be used to assess mastery of distinct subareas within a unit?

Urevbu, Andrew O. "Teaching Concepts of Energy to Nigerian Children in the 7-11 year old age Range." Journal of Research in Science Teaching, 21 (3): 255-67, 1984.

Descriptors--*Concept Formation; *Concept Teaching; *Curriculum Development; Elementary Education; *Elementary School Science; *Energy; Science Education; Science Instruction; *Scientific Concepts

Expanded abstract and analysis prepared especially for I.S.E. by Lloyd H. Barrow, University of Missouri-Columbia.

Purpose

The purpose of this study was to determine the level of attainment of Nigerian primary school children on specific energy concepts when taught at three levels - descriptive, competitive, and quantitative.

Rationale

The author cites previous work on concept learning by authorities such as Voelker (1973). All resources in the introduction section were earlier than Voelker's synthesis article.

Research Design and Procedure

The sample of the study consisted of 450 students randomly selected from primary levels (grades assessed 2-6) in one school that was randomly selected from 100 schools. The study utilized a split-plot design with treatment and level of instruction (descriptive, comparative, or quantitative) as fixed factors. ANOVA procedure was utilized to analyze the data.

The students were taught by six preservice teachers. Three different treatments were incorporated in the study - Experimental (pretest, instruction, and posttest), Control (pretest and posttest only), and Posttest (instruction and posttest only). Ss were taught four lessons focusing upon different energy concepts:

Lesson 1 sources of energy
Lesson 2 focus of energy
Lesson 3 energy derived from foods
Lesson 4 energy is the capacity to do work

The pretest was administered in a Piagetian-type interview. The pretest consisted of a 25 item test to assess concept attainment of the different level of concepts. A total of 30 minutes was provided for each interview. Osborne and Gilbert's (1980) technique was utilized to assess descriptive concepts. Subsequently, Ss were asked to supply reasons for their response. Comparative items involved explaining and predicting the order of elements/substances and their amount of energy, while quantitative items required Ss to assign number values. The pretest also was utilized as the posttest.

Findings

The posttest scores were significantly higher than the pretest scores. The researcher acknowledges that the pretest interview was instructional. Additional significant findings were superior performance by older children's interaction between grade level and pretest-posttest performance; overall differences in test performance on the descriptive, comparative, and quantitative levels of concepts; and interaction between test performance and the three levels of concepts.

Interpretation

The level of concepts were determined to be hierarchical. Performance on descriptive items were superior to comparative, which were superior to quantitative. Ss at advanced grades performed better than younger Ss on both pretest and posttest. The Piagetian explanation that older Ss can demonstrate more logical operations was inferred.

ABSTRACTOR'S ANALYSIS

The researcher attempted to assess a large group os Ss while utilizing individual interviews of 450 Ss for both pretest and posttest. However, the number of interviewers nor the competency of the interview(s) are not reported. The time interval, 25 questions within 30 minutes, doesn't provide for wait-time.

The seven energy concepts were taught at all three levels. The number of science educators and their energy expertise that validated the tests were omitted in the article. The concepts fail to include the two types of energy - potential and kinetic, but forms are included.

The utilization of six preservice teachers to teach the four lessons (one teacher/grade level) was attempted to be controlled by standardizing the instructional time/concept, team planning for teaching while utilizing similar strategies and resources. The report fails to provide background information about the six preservice teachers, class size, previous energy training of preservice teachers and Ss, and abstractness of the fourth lesson (capacity to do work). Examples of each level's assessment would allow readers to be able to assess whether Bloom's levels could be an additional variable.

The major design/analysis flaw was the failure to explain the two significant interactions; first, between year and test performance; second, level of concepts and test performance. The researcher should not acknowledge main effects when significant interactions aren't addressed. The factor analysis tables should have been reported in standard format rather than reporting only F value and level of significance.

Rudnitsky, Alan and Charles Hunt. "Children's Strategies For Discovering Cause-Effect Relationships." Journal of Research in Science Teaching, 23 (5): 451-464, 1986.

Descriptors--Cognitive Processes; *Elementary School Science; Intermediate Grades; *Learning Strategies; *Problem Solving; *Relationship; Science Education

Expanded abstract and analysis prepared especially for I.S.E. by William R. Brown, Old Dominion University.

Purpose

The purpose of the study was to describe children's problem-solving performance as they figured out how to control a dot of light on a computer monitor by pressing particular keys.

Rationale

The investigators were interested in exploring approaches or strategies used by children to solve a complex problem. Some school instruction is, or should be, aimed at teaching children to think logically or scientifically. There is little empirical support for claims that learning to write computer programs requires and enhances logical thinking. This study was designed to generate and describe some detailed data about children's thinking while using a computer. The transition between implicit, immature problem-solving strategies and explicit, mature theorizing characteristics of scientific problem-solving was explored.

Research Design and Procedure

Subjects (n=41) were told that a dot of light on a computer monitor represented a vehicle. The problem solution was to determine what effect each colored key had on the movement of the vehicle. Key strokes had to be made two or three at a time.

Two aspects of this problem varied: perspective and number of operative keys. Half the subjects received the drive perspective; they were inside the vehicle and driving it. Half the subjects received the control perspective; they were in a control tower and were operating a remote-controlled vehicle. One version of the problem used three color keys. The other version used four operative keys.

Each subject was run through the problem individually. An investigator talked with the subject during a session. All conversations were tape recorded and synchronized to the record of moves made by the subjects.

The subjects were randomly selected from two schools. All were in the fourth, fifth, or sixth grade. Laboratory school children had experience with Logo. Public school children did not have any prior computer instruction or experience.

The variables consisted of: (1) drive or control perspective, (2) number of operative keys, and (3) laboratory or public schools.

Episodes, move sequences that repeat or otherwise form a pattern, were established. Exploration was trying combination of moves rather freely. In patterns, repetition of a particular two or three key move occurred. Focusing was putting together various combinations or permutations of keys or pursuing a hunch about the effects of a particular key. The fourth episode was hypothesis testing. This involved making and testing a prediction.

The effects of the variables on each episode were analyzed using ANCOVA with subjects' age as a covariate.

Findings

Almost all subjects began the problem with exploration. Pattern move sequences were by far the most numerous type of move. Exploration and patterns sometimes led to focusing.

The laboratory school students had a greater proportion of focusing than the public school students; a greater proportion of hypothesis testing; and a lower proportion of exploration.

For focusing there was a perspective X number of keys interaction. Hypothesis testing showed a school X perspective interaction. There was more hypothesis testing in the laboratory school control perspective than in the public school control perspective. There was more hypothesis testing in the laboratory school control perspective than in the laboratory school drive perspective.

Interpretations

Exploration is a requirement for knowledge acquisition. This knowledge is necessary in order to construct a theory. The prevalence of pattern moves reflects the appreciation of the need for regularity and predictability as a necessary condition for constructing a theory. Focusing seems to mark the critical time for model or theory building in this problem. In hypothesis testing the subjects set up tests for their theories and seemed to do so intentionally.

The investigators state that theorizing is more sophisticated when it is operating on theories constructed by the subject rather than on theories or hypotheses provided by the investigator.

Subjects from the laboratory school generated more theories than those from the public school. The experiences with Logo may have helped subjects connect the current problem to relevant prior knowledge. This only occurred in the control perspective group.

The investigators conclude that subjects who are able to formulate meaningful problem representations are better able to theorize.

ABSTRACTOR'S ANALYSIS

It is clearly stated by the investigators that this is an exploratory study. The purpose of using a drive or control perspective is not clear. Both modes are abstract. The dot of light represents a vehicle. The control perspective would appear to be more in line with the real world of 9 to 11 year olds who might have experience with a remote controlled toy. The drive perspective would be totally abstract for most. The Logo turtle would certainly be similar to the control perspective. Why not use just the control perspective?

Why were some subjects given a three-key problem and others a four color key version? This abstractor is not clear as to the function of this variable. The major variable of the study appears to be the Logo user vs. non-Logo user.

The investigators stated that theorizing is more sophisticated when it is operating on theories constructed by the subject rather than on theories or hypotheses provided by the investigator. How does this generalization arise from this study? The children constructed their own hypotheses. Were any hypotheses provided to the children by the investigators?

A final comment relates to the problem-solving issue. Is directing a dot of light on a computer screen something that a subject sees as a problem to be solved? What is the connection between problem-solving and personal motivation to solve a problem? Perhaps moving a dot on a screen is an exercise that does lead from exploration to hypothesis testing. If this is correct, can these skills be transferred to a "real-life" problem that is significant to the subject?

RESPONSES TO CRITIQUES

IN RESPONSE TO THE ANALYSIS OF

Ogunniyi, M. B. "An Investigation of the Nature or Verbal Behaviours in Science Lessons" by Gene F. Craven. Investigations in Science Education, 14 (4): 26-34, 1988.

M. B. Ogunniyi
University of Ibadan

Gene Craven's comments on the article above are terse, graphic, and succinct. The study above was exploratory in nature. In 1981, when the investigation was carried out, there had been no published work in the area in this part of the world. If anything at all, it was intended to provide a different perspective to research in science teaching in Nigeria. According to Ogunniyi (1981-82):

> Most of the research studies in science education in Nigeria have been concerned mainly with the compilation of ex-post facto information... The data collected are relevant to the extent that we are made aware of what is or is not available in our schools with respect to science teaching, but they are inadequate in that we do not even know what types of activities go on in the science classrooms and/or laboratories and the consequences such activities.

It was against this background that the study in question was carried out.

But exploratory as the study has been, it has stimulated a lot of research interest in the area. The first published aspect of the study was a detailed qualitative and quantitative analysis of the types of questions used by the science teachers (Ogunniyi, 1981). The second report (i.e., the present study) submitted to Science Education in 1981 finally came out in 1984. Between 1981 and 1984 other reports examining different aspects of interaction behaviours have been published, e.g., the effect of training on teacher classroom behaviours and students' attitudes to science (Ogunniyi, 1981-1982), nature of verbal and non-verbal interactions in science (Ogunniyi, 1983). Also, several other works have been published in the area (e.g., Oludotun & Ogunniyi, 1984; Okebukola & Ogunniyi, 1984; 1986).

Several unpublished theses, here in Ibadan and other universities in the country, have examined the effects of various verbal and non-verbal interactions on students' cognitive performance in physics, chemistry and biology. Others have examined how such interactions affect process skills manipulative skills and attitudes towards science, etc. The emerging picture from these studies has been that:

1. Certain interaction behaviours are universal despite variations in the socio-cultural environments.
2. Indirect verbal interactions have been found to have a positive effect on students' attitudes towards science.
3. Context of interactions (e.g., cooperative, competitive and individualistic settings) have differential effects on students' participation, achievement and attitudes.
4. Training of the teachers is significantly related to their classroom behaviours (Ogunniyi, 1981-1982; 1983; Oludotun & Ogunniyi, 1984; Okebukola & Ogunniyi, 1984, 198(; Okebukola, 1984; etc.).

I agree with Craven's suggestion that qualitative strategies should be used in addition to quantitative analysis. In fact, last year two of my research students used such an approach in line with the suggestions made by Eggleston and others (Eggleston, et. al, 1975; Dollan & Clarke, 1979; Stubbs & Delamont, 1977; Wragg, 1973; Bellack, et al., 1960; Dunkin & Biddle, 1974; Dollan & Clarke, 1979). Their findings have certainly provided fresn and informative data in the area.

The three observers used in the study have backgrounds in physics, chemistry and biology. Besides, they were monitored by me throughout the study period to limit the problem of instrumentation. Also, there was consistency between my scores and the observers' scores throughout the study period.

Another interesting discovery during the study was that any sample of lessons observed in any subject or across the three science subjects gave the same pattern of interaction behaviours (Ogunniyi, 1983). This, as has been pointed out, is not unrelated to traditional teaching in a traditional interactional context.

No doubt, the teachers and their students were well aware of the observers. However, they knew very well that the observers were not inspectors since they had been well aware of our mission. But, in spite of this awareness and possible Hawthorne effect, their performance was still far below expectation. In other words, their best performance (artificial or otherwise) did not meet the expectation of the new curricula. The data were obtained in-class. The high agreement ratio and interrater reliability are sufficient indices of a carefully conducted investigation. Although all the lessons were recorded, only a random sample were verified afterwards. This is inevitable when it is realised that a vast amount of inf rmation is generated. The inclusion of a number of factors, viz: school, community, demography, etc., as sug ced by Craven would certainly add more clarity to findings in the area.

Whatever strategy is employed in categorizing classroom protocols, it should be borne in mind that the whole area is very complex indeed. According to Delamont and Hamilton (1977) "no single technique or theory can capture the complexity of classroom life."

Whatever causal relationship we may be tempted to establish (e.g., through inferential statistics) we must be cautious of making sweeping generalizations. The link between presage, context, and process variables on the one hand and product variables on the other is very tenuous indeed. As Dunkin and Biddle (1974) have aptly warned, the assumptions of inductive statistics are hardly met in the typical teaching-learning process. It is for the same reason that qualitative analysis as suggested by Craven seems apposite but even this is beset by a different array of problems. The concept of classroom interaction behaviors and the methods of analysis cut across various fields, viz: social psychology, anthropology, sociology, sociolinguistics, communication arts, and even philosophy. The problem of subjectivity is not easy to overcome - but someone might say problems are there to be solved! The essentially a priori reductionism inherent in quantitative interaction analysis has its merits and demerits. In the same way the holistic, unsystematic and open-ended framework of ethnographic research has its merits and demerits. A combination of both methods might produce a more comprehensive picture of classroom interaction behaviors although the underpinning assumptions are somehow different.

REFERENCES

Bellack, A. A., H. M. Kliebard, and W. J. Elena (eds.). The Language of the Classroom. New York Teachers College Press, Columbia University, N. Y., 1966.

Delamont, S. and D. Hamilton. "Classroom Research: A Critique and A New Approach," in Stubbs, M. & S. Delamont, 1977, op. cit. p. 3.

Dollan, T. and P. A. Clarke. "Language in Science Lessons." The School Science Review, 61 (215): 342-344, 1979.

Dunkin, M. J. and B. J. Biddle. The Study of Teaching. Holt, Rinehart and Winston, Inc., N. Y., 1974.

Eggleston, J. F., M. J. Galton, and M. E. Jones. Final Report of the Schools Council Project for the Evaluation of Science Teaching. Open University Press, Milton Keynes, U. K., 1975.

Ogunniyi, M. B. "Classroom Interactions Relative to Teacher-Student Question Behaviours." Journal of the Science Teachers' Association of Nigeria, 19 (2): 116-122, 1981.

Ogunniyi, M. B. "An Exploratory Study of the Nature of Science Teaching in Selected Nigerian Secondary Schools." African Journal of Educational Research, 4 (1 & 2): 81-96, 1981-82.

Ogunniyi, M. B. "An Analysis of Laboratory Activities in Selected Nigerian Secondary Schools." European Journal of Science Education, 5 (2): 195-201, 1983.

Okebukola, P. A. O. "In Search of A More Effective Interaction Patterns in Biological Laboratories." Journal of Biological Education, 18 (4): 305-308, 1984.

Okebukola, P. A. O. and M. B. Ogunniyi. "Cooperative, Competitive and Individualistic Interaction Pattern: Effects on Students' Achievement and Acquisition of Practical Skills." Journal of Research in Science Teaching, 21 (9): 875-884, 1984.

Okebukola, P. A. O. and M. B. Ogunniyi. "Effects of Teachers' Verbal Exposition on Students' Level of Class Participation and Achievement in Biology." Science Education, 70 (1): 45-51, 1986.

Oludotun, J. S. O. and M. B. Ogunniyi. "The Performance of Selected Physics Teachers Relative to the Objective of C.E.S.A.C." Journal of the Curriculum Organization of Nigeria, 2 (2): 20-25, 1984.

Stubbs, M. and S. Delamont. Explorations in Classroom Observations. John Wiley & Sons, Chichester, U. K., 1977.

Wragg, E. C. A Study of Student Teachers in the Classroom. The Open University Press, Milton Keynes, U. K., 1973.

IN RESPONSE TO THE ANALYSIS OF

Stanley, Julian C. and Barbara S. K. Stanley. "High-School Biology, Chemistry, or Physics Learned Well in Three Weeks" by Constance M. Perry. Investigations in Science Education, 14 (4): 56-63, 1988.

Barbara S. K. Stanley
Julian C. Stanley
Johns Hopkins University

Professor Perry has prepared a magnificent expanded abstract and analysis of our article. We greatly appreciate the time, energy, and thought she devoted to it. Let us try to put our study into context before responding to her specific points.

The Study of Mathematically Precocious Youth (SMPY) was founded by Julian Stanley at Johns Hopkins in 1971 to find youths who reason exceptionally well mathematically and provide them the special, supplemental educational opportunities most of them sorely need and richly deserve. It has led to large talent searches and extensive academic programs at Johns Hopkins, Duke University, Northwestern University, the University of Denver, Arizona State University, Iowa State University, Sacramento State University, the University of Wisconsin at Eau Claire, the University o. Washington in Seattle, the States of Illinois and Minnesota, and elsewhere. At least 100,000 boys and girls less than 14 years old take the College Board Scholastic Aptitude Test in a given year (usually in January) as entrants in a regional or state talent search. When SMPY began, probably less than a dozen did.

Also, there is now an SMPY at Iowa State University" and an "SMPY at Tianjin, People's Republic of China." In the fall of 1989 Dr. Ann E. Lupkowski will found "SMPY at the University of North Texas" in Denton. From the vast experience with mathematically talented youths that Dr. Stanley and his associates have had over the years, a number of their suggestions and recommendations are derived. SMPY is administratively independent of all the above-mentioned organizations, but of course benefits greatly from their experimentation.

Specific Comments

First, there are two errors on page 242 of our article, which Professor Perry did not detect: the last two lines of the text should read "compared with the 56% (not 70%) of the biology class who took the equivalent biology test and averaged 4.36 (not 4.42)!" Apparently, they occurred because of inadequate data in early drafts of the article. The correct figures can be computed from data on page 241, where the Advanced Placement examination scores of the 14 biology-class students (out of 25) who took the exam are recorded. We regret making these errors. of course, but note that they affect our conclusions little.

The "unbiased" estimate of the standard deviation of the chemistry posttest scores in 1982 was 62. Those scores ranged from two 800s (above the 99th %ile) to one 600 (59th %ile). No chemistry pretest was administered by us in 1982 because one of us, a former high school chemistry teacher, suspected that even extremely bright youths were not likely to have much technical knowledge of chemistry before taking a course in it. The low percentile ranks of most chemistry pretest scores in later years bore out this expectation. (The score scale of one achievement test is not directly comparable with that of another.)

Our chief intent in creating the three-week science courses was to free a student's time in school the following fall so that he or she might go directly into the second year of the subject, at the College Board Advanced Placement Program level. There are usually, at most, five high school science and mathematics courses at a second-year level: biology, calculus, chemistry, computer science, and physics. If the elementary level of each had to be taken first as a regular school course, this would amount to as many as 10 subjects, virtually an impossibility to schedule and carry. By completing one or two courses well in a single summer (four of the 34 students in 1982 distinguished themselves in both biology and chemistry), the student can move on to the next level that fall.

It was our observation, confirmed by the students' questionnaire responses, that most of the students in biology enjoyed that class more than any science class they had ever taken in school. The chemistry students were less enthusiastic, probably because their teacher seemed not nearly as well informed, stimulating, or enthusiastic as the biology instructor. Several students said they had signed up for biology in order to "get it out of the way" because they expected to have little interest in the subject. One of these earned the highest possible score, 800, on the biology posttest and changed his attitude toward the subject markedly.

Note, however, that on the achievement posttest the biology students' median score was at the 95th percentile of the highly select national group that elects to take the test, and the chemistry students' median had exactly the same percentile rank, 95, with respect to the chemistry national-norm group. Students seem to have learned the subject matter at the high school level equally well in both classes, but with different zest and desire to continue in the field.

We are now in the process of following up the 1982 groups to see how they have fared educationally since then. Also, Dr. Sharon J. Lynch, formerly of the Johns Hopkins University Center for the Advancement of Academically Talented Youth (CTY), has prepared an extension of our study, entitled "Fast-paced High School Science for the Academically Talented: A Six-Year Perspective." It is being considered for publication by a professional journal. Her comparisons go from 1982 through 1987, year by year. Interested persons may write to her c/o Peter Perenyi, AM-CON-GEN (POZ), APO, New York 90213. Students who get this early start into the sciences can readily take more Advanced Placement Program examinations by the time they complete high school than most other students. Those AP credits may help them gain admission to more-highly-selective colleges and greater curricular flexibility there.

Unanswered Questions

We cannot "know how these same or comparable students would do on the same tests after taking a traditional 5-day-a-week, year-long course," because it would be extremely difficult even to devise a study bearing on that question that involved randomized assignment to the experimental group vs. a control group. Students electing to take a regular high school science course in three weeks are a very special group of volunteers for intensive study of a particular subject. Their stupendous achievement relative to the norms of extremely difficult achievement tests (e.g., 100 items in 60 minutes, designed for those high school juniors or seniors with the greatest background and interest in the subject) suggests strongly, however, that they would not likely have done as well after the usual school year of the subject.

Specific Comments

First, there are two errors on page 242 of our article, which Professor Perry did not detect: the last two lines of the text should read "compared with the 56% (not 70%) of the biology class who took the equivalent biology test and averaged 4.36 (not 4.42)!" Apparently, they occurred because of inadequate data in early drafts of the article. The correct figures can be computed from data on page 241, where the Advanced Placement examination scores of the 14 biology-class students (out of 25) who took the exam are recorded. We regret making these errors, of course, but note that they affect our conclusions little.

The "unbiased" estimate of the standard deviation of the chemistry posttest scores in 1982 was 62. Those scores ranged from two 800s (above the 99th %ile) to one 600 (59th %ile). No chemistry pretest was administered by us in 1982 because one of us, a former high school chemistry teacher, suspected that even extremely bright youths were not likely to have much technical knowledge of chemistry before taking a course in it. The low percentile ranks of most chemistry pretest scores in later years bore out this expectation. (The score scale of one achievement test is not directly comparable with that of another.)

Our chief intent in creating the three-week science courses was to free a student's time in school the following fall so that he or she might go directly into the second year of the subject, at the College Board Advanced Placement Program level. There are usually, at most, five high school science and mathematics courses at a second-year level: biology, calculus, chemistry, computer science, and physics. If the elementary level of each had to be taken first as a regular school course, this would amount to as many as 10 subjects, virtually an impossibility to schedule and carry. By completing one or two courses well in a single summer (four of the 34 students in 1982 distinguished themselves in both biology and chemistry), the student can move on to the next level that fall.

It was our observation, confirmed by the students' questionnaire responses, that most of the students in biology enjoyed that class more than any science class they had ever taken in school. The chemistry students were less enthusiastic, probably because their teacher seemed not nearly as well informed, stimulating, or enthusiastic as the biology instructor. Several students said they had signed up for biology in order to "get it out of the way" because they expected to have little interest in the subject. One of these earned the highest possible score, 800, on the biology posttest and changed his attitude toward the subject markedly.

This is consistent with SMPY's rather vast experience in mathematics. For example, in one of our early programs students who had completed a year of first-year algebra in school were found to know less algebra than younger students who had studied it only 20 hours. Furthermore, the former had formed habits of inattention that interfered greatly with their learning second-year algebra fast, whereas the latter had no difficulty doing so.

" . . . a clarification as to why the three-week course is deemed more stimulating would be beneficial." We believe that is mainly because all of these extremely able students are fully capable of moving ahead fast and well at a high level of rigor, and are eager to do so. They are not held back by slow learners. Also, all are excellent reasoners, so the teacher does not have to spend considerable time on the reasoning aspects of the subject, which in typical classes must be done. Most are intensely interested in the subject, which is why they chose the rigors of the course in preference to any other thing they might have done during those three weeks.

No, science is not always "boring for the intellectually capable student if taken in a traditional year-long program." Yes, there are a few "high school science teachers who can work with such students and not bore them." They, despite grueling teaching and activities loads, can inspire, stimulate, and motivate even the ablest students all the way to the Westinghouse Science Talent Search and beyond. We suspect that few students who have the prospect of working with such super-teachers will enroll for a three-week summer course in their subject, instead.

From observing teachers of many different types of fast-paced courses from 1972 to the present, we have concluded that thorough knowledge of the subject is the sina qua non, the prime condition, necessary but not sufficient in itself. Zeal for the subject, liking intellectually talented youths, not feeling threatened by them, and a number of hard-to-define qualities are also essential. When a group of intellectually talented volunteers is stimulated by a masterful fast-pacing teacher, good things happen for most of the students. The process is remarkably robust, given that both the students and the teacher are highly able and eager to cover that particular subject quickly and well. Questionnaire responses from students and erviews with them over the years have strengthened these surmises.

SMPY started the classes in 1982 to test our belief that they would succeed. CTY has continued them each year thereafter, adding physics in 1983, but has not systematically studied the students' reactions in quite the detail that we did. In the article, test results in 1984 were compared with those in 1982 to see whether or not sizable gains were still being made two years later.

Laboratory time was greatly increased in 1983 and thereafter. Median gains from pretest to posttest in biology were 167 points in 1982, and 124 and 197 (average of about 160) in 1982. Thus, from this standpoint, the decrease in class time seems not to have hurt achievement.

SAT-V Score Important?

SAT-Verbal scores of a few students were in the middle 400s. Such persons tended to be among the lower scorers on the biology posttest, but even the lowest one of the 25 scorers ranked better than 61% of the highly able norm group. Relatively low verbal ability seemed essentially unrelated to posttest scores in chemistry. The lowest SAT-V scorer (340V, 750M at age 12) ranked last on the biology posttest but later did well in architecture at a famous university.

The lowest ranker on the chemistry posttest (510V, 510M at age 12) is doing well at a great university and planning to continue toward a Ph.D. degree in theoretical physics. As a college sophomore (1988-89), he sent SMPY the following comments: "The chemistry course . . . sparked my interest in investigation. I had always been interested in science; CTY showed me it need not be watered down. This was the push which enticed me into a career in science. . . . I still enjoy chemistry and may pursue it as a minor. . . . The course was my first escape from the explanations of general science courses, which just stated [that] a phenomenon happens, to explanations of <u>why</u> the event happens." (Most higher scorers on the chemistry test had not been as favorable on the post-course questionnaire as he was even then.)

Longer-Term Effects of Courses

Although it is difficult to prove, we believe that "intellectually highly able students who move through science coursework more rapidly than is traditional [are] more likely to pursue higher degrees and careers in science than their counterparts who follow a traditional timetable." Of course, the eagerness, ability, motivation, and interest in science indicated by voluntary enrollment in the three-week courses marks these students as being especially promising.

Anecdotal evidence strengthens our conjecture. For example the 13-year-old who ranked next to the bottom of the biology class on the posttest later majored in biology at one of the country's greatest universities and graduated in seven semesters at age 19 with an all-A record. (Her SAT scores at age 12 were 490V and 700M.) A professor and dean there wrote the following about her: "[She] is one of the most impressive (perhaps the most impressive candidates for medical school or graduate school I have known in my 34 years of teaching at . . ."

A high scorer on the biology posttest (750; 570V, 570M at age 13) comments as following in 1988: Junior at a famous college, working toward a "Ph.D. in molecular biology . . . Without the biology course at CTY, I would probably be a computer science major. The teacher . . . was one of the best I have ever met . . . The broad scope of the course allows me to follow developments in all aspects of biology . . . with ease and lets me see theories in their larger context."

Also, the caliber of schools now being attended by these former students is most impressive. For the first 21 respondents to the present follow-up they are as follows: U. California at Berkeley (then graduate work at U. Pennsylvania), Caltech (then Purdue), Case Western Reserve (then Cornell), Cornell, Dartmouth, U. Delaware, Duke, Harvard (2), Johns Hopkins (then Medical College of Virginia), MIT (4), Princeton (3), U. Richmond, U. Virginia, U. Washington at Seattle (then U. California at Berkeley), and Yale.

The systematic follow-up study of the 1982 classes now being conducted at Johns Hopkins should provide better evidence about the college- and, in some cases, graduate-school success of these pioneering students.

Short-term, intensive, academic summer courses are an option that should be available to those highly able youths who want it. Social benefits of interacting closely with one's intellectual peers make the residential setting especially attractive and beneficial. Perhaps the chief advantage of successfully completing such courses is increase in self-confidence. Entering the most selective colleges and universities in the land several years later will hold far fewer terrors for them than for most students who have attended only high school.

"The report was, however, somewhat choppy and therefore confusing. . . ." Yes, even we had some difficulty finding certain statements and figures in it these several years later. A more time-linear progression might have made points in the article easier to locate and keep in mind.

Educators who wish to observe these classes at Franklin and Marshall College in Lancaster, Pennsylvania, where two consecutive three-week terms of each subject are conducted by CTY every summer, should write to Dr. Luciano Corazza, CTY, 2933 N. Charles Street, Baltimore, Maryland 21218, or telephone him at (301) 338-8427.

END

U.S. Dept. of Education

Office of Education
Research and
Improvement (OERI)

ERIC

Date Filmed

March 21,1991