DOCUMENT RESUME

ED 168 882 SE 027 510

AUTHOR Blosser, Patricia E., Ed.; Steiner, Robert L., Ed.
TITLE Investigations in Science Education, Vol. 5, No. 2. Expanded Abstracts and Critical Analyses of Recent Research.
INSTITUTION Ohio State Univ., Columbus. Center for Science and Mathematics Education.
PUB DATE 79
NOTE 66p.
AVAILABLE FROM Information Reference Center (ERIC/IRC), The Ohio State University, 1200 Chambers Rd., 3rd Floor, Columbus, Ohio 43212 (subscription $6.00, $1.75 single copy)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Achievement; *Attitudes; *Educational Research; Inservice Teacher Education; Instruction; Preservice Education; Research; *Research Methodology; *Science Education; Teacher Education
IDENTIFIERS *Instrument Development

ABSTRACT

This issue of "Investigations in Science Education" contains abstracts and critical analyses of eight articles grouped into three topical areas: (1) instrument development; (2) attitude studies; and (3) achievement. In addition, it also contains a response to the analysis of an article included in a prior issue. Each abstract-analysis includes biographical data, research design and procedure, purpose, research rationale, and the abstractor's analysis of the research. (PEB)

INVESTIGATIONS IN SCIENCE EDUCATION

National Association for Research in Science Teaching

ERIC Clearinghouse for Science, Mathematics, and Environmental Education

Published Quarterly by

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

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

Volume 5, Number 2, 1979

NOTES FROM THE EDITORS

With this issue of INVESTIGATIONS IN SCIENCE EDUCATION (Volume 5, Issue 2), we welcome two new members to the Advisory Board: Willard J. Jacobson, who is an ERIC-SMEAC appointee, and Livingston S. Schneider, a NARST appointee. We also want to thank the two persons retiring from the Advisory Board; David Butts and Kenneth Jacknicke, for their assistance during the past several years.

Issue 2 of Volume 5 contains eight studies clustered in three areas: instrument development, attitudes, and achievement. It also contains a response from an individual whose study was critiqued in a previous issue of I.S.E.(Volume 3, Issue 1). We are pleased that authors are responding to our offer to allow them to reply to questions, requests for clarification, or criticisms raised by I.S.E. abstractors. We hope the dialogue will continue.

Patricia E. Blosser
Editor

Robert L. Steiner
Associate Editor

INSTRUMENT DEVELOPMENT

Kozlow, M. James and Marshall A. Nay. "An Approach to Measuring Scientific Attitudes." Science Education, 60(2):147-172, 1976.
Descriptors--*Attitudes; *Attitude Tests; Educational Research; Evaluation; Science Education; Secondary Education; *Secondary School Science; *Scientific Attitudes; *Test Construction

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Robert L. Shrigley, The Pennsylvania State University.

Purpose

The purpose of this study was to design and field-test a two-part instrument having cognitive and affective components that would measure scientific attitudes.

Rationale

The steps to designing the instrument were:

1. Review and cite the shortcomings of five instruments:
   A) Test on Understanding Science
   B) Attitudes Toward Science and Scientific Careers
   C) Projective Test of Attitudes
   D) Science Support Scale
   E) An Inventory of Scientific Attitudes

2. Citing the difficulty other researchers had applying the Krathwohl taxonomy of affective objectives to the natural sciences, the authors chose instead the Nay-Crocker model. From the 65 affective attributes of scientists included in the Nay-Crocker inventory, the authors chose eight scientific attributes for their instrument: 1) critical-mindedness; 2) suspended judgment; 3) respect for evidence; 4) honesty; 5) objectivity; 6) willingness to change opinions; 7) open-mindedness; and 8) questioning attitude.

3. Design the Test of Scientific Attitude (TOSA) on the three dimensions of attitude suggested by Rokeach (cognition, affection [intent] and behavior [action]).

A) Cognition: tests the students' understanding of the manner in which scientists manifest the scientific attitude (resulting in the 20-statement Cognitive Component Subtest-CCS).

B) Intent: tests the students' tendency to approve or disapprove specific courses of action representing elements of the scientific attitude (resulting in the 20-statement Intent Component Subtest-ICS).

C) Action: represents the demonstrative behavior of students in the classroom implying a scientific attitude (with each student to be judged by science teachers on a four-point scale).

4. A multiple choice format providing the student with four courses of action was chosen for TOSA after documenting weaknesses in Thurstone's equal-appearing interval, Likert's summated rating, and Osgood's semantic differential techniques.

Procedures

Using Phases I and III of Engman's four-phase model for educational planning, the authors wrote one general and four or five specific behavioral objectives for each of the eight Nay and Crocker scientific attributes. The objectives were submitted to judges who rated each of them on a 0-2 scale of relevance. As a result the initial list of objectives was reduced to less than half. Openmindedness was subsumed under objectivity and the questioning attitude was subsumed under critical-mindedness, reducing the eight original categories of scientific attributes by two.

Twenty-eight multiple choice items representing the six categories were then constructed and field-tested. As a result of a pilot study, some of the 28 items were revised and 12 new ones added to complete TOSA, 20 items each for CCS and ICS. A panel of judges reacted to the keyed response for each test item.

Findings

When the instrument was administered to 307 secondary school physics and chemistry students the means for TOSA, CCS and ICS were 52.4, 52.8 and 52.2 percent, respectively, and the standard deviations were 10.3, 13.9 and 12.5.

Item Analysis. None of the alternatives on the 39-multiple choice items (item 20 was dropped due to a misprint) was ignored by all students, but seven of the 156 distractors accounted for less than 3 percent of the responses. Twelve of the 40 items were outside the desired .25-.75 difficulty range. Two of the 39 items fell well below the desired .30 on the biserial correlations.

Reliability. The correlation coefficient between CCS and ICS was .23, well below the .40 odd-even, split-half correlation of TOSA, implying that the subtests were measuring different entities. Testing for homogeneity, the KR-20 coefficients for TOSA, CCA and ICS were .55, .45, and .39, respectively. Testing for instrument stability, TOSA was administered to 105 students during a three-week, test-retest plan. The correlation coefficients for TOSA, CCS and ICS were .71, .68, and .64, respectively. The correlation coefficients for TOSA, CCS and ICS with reading as indicated in the scores of 248 students on the Sequential Test of Educational Progress were .35, .41, and .18, respectively. This implied less than a strong relationship between reading ability and attitude scores for physics and chemistry students.

Validity. Content validity was assumed as a result of the juror involvement and the rigorous procedure whereby content was selected for TOSA. Structural validity was examined through factor analysis with 35 of the 39 items loading on nine factors. The factor analysis provided reasonable support for TOSA's six classifications of scientific attitudes as originally categorized by the authors.

As a means of establishing external validity of TOSA, three science teachers were asked to rate the scientific attitude of their students (the n = 151 of total 307) on a four-point scale. The correlations of teacher ratings with student scores on TOSA, CCS and ICS were not only low, but inconsistent from teacher to teacher. As a result, the authors discarded teacher ratings as a procedure for establishing external validity of their attitude instrument.

Action Component. By asking teachers to rate the scientific attitude of students, the authors attempted to satisfy the action component of the attitude concept. Because the correlations between teacher ratings and student scores were low, the authors suggest that action is not only dependent on a prestated attitude, but also the situation.

Conclusions.

1. The factor analysis supported the author's rationale for dividing the test items into CCS and ICS and also their classification on the basis of behavioral definitions of the attitudes.

2. CCS (cognitive) and ICS (intent) are not measuring the same characteristics.

3. Understanding the attributes of the scientists will probably not ensure that students will demonstrate those characteristics.

4. Using the analysis data, test items should be modified and instrument lengthened.

5. The action component of attitude needs more attention.

6. The generalizability should be broadened by submitting the instruments to a wider population of students.

ABSTRACTOR'S ANALYSIS

1. TOSA is obviously developmental and the authors were adequately modest in their appraisal of their professional accomplishments.

2. When compared with other instruments measuring scientific attitude, the authors' major contributions were an instrument:

(1) with a more rigorous step-by-step rationale, (2) with a unique multiple choice format, and (3) representing science classroom situations and experiences.

3. The authors recognized the importance of the action component of the attitude concept, a facet of attitude research in science education which is usually ignored. The authors confirmed a finding so common in attitude research. The literature is strewn with studies that seem to deny a strong relationship between attitude and behavior.

Fishbein and Ajzen (1975) may have come to grips with this problem. They question the assumption that there is a one-to-one relationship between attitude and behavior, the former representing the stimulus and the latter a response. Instead of a S-R relationship, they offer the alternative that there is a probabilistic relationship.

Because of situational factors, or what Fishbein and Ajzen call subjective norms (pressure from significant referents), a subject whose attitude is positive would respond positively to many, but not necessarily all, manifestations of a particular psychological object.

For this reason Fishbein and Ajzen suggest that behavior cannot be measured by a single observation but rather by repeated observations. Kozlow and Nay do not report the exact procedure expected of the teachers when they rated the scientific behavior of the students.

4. Central to the authors' rationale was the development of an instrument centered around situations created in the students' science classroom. And this is probably a worthy objective. So their decision to reject the Osgood semantic differential attitude measuring technique with its unique bipolar adjectives format (e.g., good-bad) was probably justified.

The authors rejected the Thurstone format on the basis that unidimensionality is assumed and their goal was an instrument representing many dimensions in the affective domain. Even more questionable may have been Thurstone's time-consuming and jury-biased validation procedures for determining the negative-positive intensity of attitude statements (Hovland and Sherif, 1952). Although the authors of TOSA used a different, and probably a less sophisticated, jury procedure, it is possible that their system for determining the "best" answer for each multiple choice item has a jury-biased problem not unlike that of the Thurstone technique.

The authors rejected the Likert procedure because of the response biased tendency in their testing formats. The abstractor recognizes the response bias in non-multiple choice testing format (Oppenheim, 1966), but the socio-psychological literature of the last two decades fails to recognize it as a major concern in attitude testing.

The multiple choice format chosen for TOSA by the authors ignores attitude intensity. In responding to each of the 40 statements, students could choose either the correct alternative or one of the three incorrect alternatives. There was no way a student could register intensity of attitude on each statement as is the case on Thurstone's 11-point, Osgood's 7-point, and Likert's 5-point positive-negative continuum.

The abstractor suggests that the Likert format with an equal number of negative and positive statements sequenced randomly probably has less handicaps as an attitude measuring scale, especially ICS, than the multiple choice format.

5. The statements on ICS seem to be the application of standard operating procedures of the scientific community within the milieu of the science classroom, thus they might be considered cognitive, even factual in nature. And one of Edward's (1957) 14 criteria for judging the validity of attitude statements is that they be nonfactual.

Although the authors' correlation of scores on ICS and CCS resulted in a low coefficient of .23 implying that ICS may not be highly cognitive, many of the ICS statements seem to lack the like-dislike polarity (Zimbardo and Ebbesen, 1969), that emotionally-charged characteristic of attitudes (Triandis, 1971), or the evaluative component that separates the concept of attitude from other related psychological phenomena (Fishbein and Ajzen, 1975).

The abstractor suggests that a close examination of other instruments said to measure scientific attitude, even those designed around a Likert format, seem to lack some of the characteristics of attitude as defined above. Therefore, a more fundamental question is involved. Science educators should consider the possibility that a scientific attitude rooted firmly in those attributes of the scientist may be inherently more cognitive than the concept of attitude as defined by the social psychologist. Even a favorable or unfavorable attitude toward the subject of science, the learning of science or the teaching of science may not be the same concept nor might it have a significant relationship with the scientific attitude as it was defined in this study.

Therefore, until the concepts, scientific attitude and attitude, are found to be closely related suggesting that such instruments as TOSA be designed from a rationale compatible with the findings of the social psychologist, science educators have to suspend any judgment of the authors' unconventional multiple choice format and the seeming cognitive nature of ICS.

REFERENCES

Edwards, A. L. Techniques of Attitude Scale Construction. New York: Appleton-Century-Crofts, 1957.

Fishbein M. and I. Ajzen. Belief, Attitude, Intention and Behavior. Reading, Mass: Addison-Wesley, 1975.

Hovland, C. I. and M. Sherif. "Judgmental Phenomena and Scales of Attitude Measurement: Item Displacement in Thurstone Scales." Journal of Abnormal and Social Psychology, 42:215-239, 1952.

Oppenheim, A. N. Questionnaire Design and Attitude Measurement. New York: Basic Books, 1966.

Triandis, H. C. Attitude and Attitude Change. New York: John Wiley, 1971.

Zimbardo, P. and E. Ebbesen. Influencing Attitudes and Changing Behavior. Reading, Mass; Addison-Wesley, 1969.

Munby, A. H.; R. J. Kitto; and R. J. Wilson. "Validating Constructs in Science Education Research: The Construct 'View of Science'." Science Education, 60(3):313-321, July-Sept., 1976.

Descriptors--Authoritarianism; Dogmatism; *Educational Research; *Evaluation; *Models; *Research Methodology; *Science Education; Test Validity; *Validity

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Rodney L. Doran, State University of New York at Buffalo.

Purpose

This study had two purposes:

(1) to demonstrate a "methodology for validating instruments employed in science education research" and

(2) to provide "empirical support to the conceptual linkage between the philosophical view of science and the intellectual variables, dogmatism and authoritarianism."

Rationale

This study was based on conclusions from several reviews of science education research that (1) many studies rely on author-constructed instruments whose validity and reliability has been established only cursorily, if at all, and (2) much research is conducted without a coherent conceptual or theoretical framework.

This research was an attempt to demonstrate that theoretically-based, psychometrically sound validation methodologies can be used. The instrument to be empirically validated was a classroom observation system developed by one of the investigators in his doctoral dissertation (Munby, 1973).

Research Design and Procedure

The multitrait-multimethod model of Campbell and Fiske (1959) was used to guide the selection of instruments and the formulation of hypotheses.

The following figure displays the written and observation type instruments that separately assessed the construct in question—"View of Science," a convergent construct—"Authoritarianism," and a divergent construct—"Performance Anxiety." Convergent measures are those deemed to have a "defined conceptual coherence" with the trait in question. Similarly, divergent measures are those that are "conceptually understood to have no relationship" to traits A and B.

	CONSTRUCT		
Method	A View of Science	B Authoritarianism (convergent)	C Performance Anxiety (divergent)
Observation	1 Munby System	3 Flanders System	5 Behavioral Checklist
Written	2 Nature of Science Scale (NOSS)	4 California - F Dogmatism - D Dogmatism - E	6 PRCS

The NOSS was chosen because it had "philosophical bases similar to that of the Munby system" and therefore was deemed to measure the same trait, using a different method. Through an analysis of the realism-instrumental framework of the Munby system, it appeared that realism reflects an "authoritarian and dogmatic view about truth and about science." Thus, the convergent observation method chosen was the Flanders interaction analysis system in which low I/D ratios are "associated with dogmatic and authoritarian teaching styles."

The divergent trait selected for use was "performance anxiety"—one of the few constructs for which more than one method of measure was available. This trait can be measured by a "written personal report of confidence as a speaker (PRCS) and a timed behavioral checklist for performance anxiety in speaking." There appeared to be no theoretical reason for this construct "...to be related to view of science or any other construct chosen for this validation study...."

The need for this present study surfaced as there were "...three equally attractive measures available for the written assessment of the convergent trait (California-F, Dogmatism-D, and Dogmatism-E)." This study was undertaken to assist in the selection of an instrument for measuring authoritarianism (the convergent trait) via a written format. Accordingly, the following hypotheses were formulated:

(1) Scores on the Nature of Science Scale would correlate positively and significantly with those on the California-F, Dogmatism-D, and Dogmatism-E scales.

(2) Scores on the PRCS would not correlate with scores on the California-F, Dogmatism-D, and Dogmatism-E scales.

(3) Random numbers added to the data analysis would show correlatives similar to those for PRCS, except that they would have near-zero reliability.

To simplify the administration and scoring of the instruments, a single test was developed which was composed of items from the "written" instruments arranged in random order. Also, the response format for the NOSS and the PRCS was changed from True-False to a Likert-type scale (with responses +3, +2, +1; -1, -2, -3) to correspond with that used in the California-F and Dogmatism scales.

Additionally, a "trait" of random numbers was included in the analysis. "Since the divergent measure performs essentially the same function, theoretically, as collecting random numbers, it was considered potentially useful to amplify the validation model...." with the random number trait of "...approximately normal distribution, corresponding to the scoring minima and maxima of the real instruments."

The 154 items of the five written instruments were administered to "168 secondary school teachers of a variety of subject specializations."

The participants were told that the purpose of the research was to validate an instrument, but they were not told the nature of variables and instruments.

From the response data, scores for the five tests were computed for each of the participants. Descriptive test statistics and reliabilities (split-half with Spearman-Brown correction) were obtained for each instrument and correlation coefficients among the five scales and the "random number" trait were calculated.

Findings

The following data were the bases for hypothesis testing and resultant conclusions.

	NOSS	F	E	D	PRCS	Random
California-F	.51*					
Dogmatism-E	.42*	.78*				
Dogmatism-D	.39*	.78*	.97*			
PRCS	.04	.09	.14	.14		
Random Number	.08	.00	.06	.06	.04	
Number of Items	29	29	40	66	30	—
Reliability	.67	.84	.83	.90	.94	.16

*Significant at .01 level.

The reliabilities of the instruments obtained in this study fall within the ranges reported in the literature. It appears the pooling of the items into one composite test did not unduly affect the characteristics of these instruments.

The convergence of the "view of science" construct (as measured by the NOSS) with the authoritarian and dogmatism construct (measured by the

D, E, and F scales) was clearly demonstrated by the significant correlation coefficients among these four variables. Accordingly Hypothesis 1 was accepted.

Similarly, Hypothesis 2 was accepted as the relationships between the PRCS and the D, E, and F scales were not significantly different from zero. Also, Hypothesis 3 was accepted as the relationships of the convergent measures (D, E, and F) with the "random number" trait were not significantly different than with the PRCS.

Interpretations

The authors concluded that the California-F scale was the most useful of the convergent measures, because: (1) it correlated best with NOSS and (2) correlated the least with PRCS.

Specifically, based on the evidence described, the authors concluded that there was support for the "...theoretical position that the view of science construct converges with authoritarianism and dogmatism, while it diverges from performance anxiety." Based on these tentative findings, the investigators will use the California-F scale and conduct the validation via the full model (all six boxes) as outlined earlier.

More generally, the authors state that this study has "...shown the potential usefulness of the Campbell and Fiske multitrait-multimethod validation model for science education research and has illustrated the power of this model...." Further, that the "procedures adopted in this research demonstrate the importance of proceeding from a sound theoretical framework and of employing a conceptually sensible and rigorous validation methodology."

ABSTRACTOR'S ANALYSIS

This study attacked two persistent problems for researchers in science education. The report communicates clearly the intent, procedures and outcomes of the research. As this appears to be the first study of its kind, the main empirical study cited was a doctoral dissertation by one of the authors. This doctoral research was apparently the seed from which this study grew.

The major methodological contribution of this study was the use of the Campbell and Fiske model from social sciences into a specific education domain. The relevant dimensions and characteristics of the Campbell and Fiske model were described in a brief but sufficient manner. This application of the model to science education research appears to be both innovative and appropriate.

The construct "view of science" basic to this study was defined as representing "a basic philosophic posture toward science and reality." The diametrically opposing views "realism," and "instrumentalism" as described by Nagel (1961) were the cornerstone of the research. The Munby classroom observation system was illustrated via nine statements consistent with each of these "views." One should consider Nagel's reservations about over-interpretations of these stances:

> It is therefore difficult to escape the conclusion that when the two apparently opposing views on the cognitive status of theories are each stated with some circumspection, each can assimilate into its formulations not only the facts concerning the primary subject matter explored by experimental inquiry but also the relevant facts concerning the logic and procedure of science. In brief, the opposition between these views is a conflict over preferred modes of speech (p. 152).

Apparently, these preferences in the way the process and products of sciencing are communicated is reflective of one's philosophical beliefs about science. According to Kimball (1967-68), author of the NOSS:

> Philosophical considerations characterize thoughtful discussion of the nature of science, so philosophy majors might

display a better understanding of the nature of science than science majors (p. 111).

It appears knowledge of specific terms and language used in discussing statements about science may be a function of academic background, thereby clouding the assessment of one's "view of science." This concern is heightened by Kimball's findings that philosophy majors surprisingly did score significantly higher on the NOSS than did science majors (higher scores indicating greater agreement with the model of the nature of science). Responses to seven items describing the "methods of science" were the main source of the overall NOSS differences. According to Kimball, "This is a remarkable outcome, for it would seem reasonable that the one area of science with which the scientists would be most familiar is methodology." This raises the question as to whether the NOSS is measuring understanding of nature of science or understanding of the unique philosophical interpretations of nature of science or understanding of the unique interpretations of certain words and phrases by these schooled in philosophy.

The authors listed the nine factors of the realist view and, similarly, the nine factors for the instrumentalist view as operationalized in the Munby classroom observation system, which wasn't used in the study reported here. The D, E, and F scales were described as being measures of authoritarianism and dogmatism. These traits were characterized by Rokeach (1960) as associated with "...concrete thinking, intolerance of ambiguity, and premature closure of perceptual process." The D and E forms of the Rokeach Dogmatism scale are indicated as the fourth and fifth revision. Further, they are described as Dogmatism and Enthrocentrism scales, respectively. However, no illustrative items were included; rather, psychometric data were listed for each, e.g. means and reliabilities. The reliability data for D Scale was described by Munby et al. of a sample of "English majors," while it was "students of University College in London," according to Robinson and Shaver (1973).

From the report it appeared that the California-F scale was developed by Rokeach. However, it was designed by Adorno and colleagues (1950) to measure "...antisemitism and ethno-centrism without mentioning minority groups or current political-economic issues..." (p. 223). "The new instrument was termed the F scale, to signify its concern with implicit prefascist tendencies..." (p. 224). Which of the three forms of the F-scale was used in this study was not stated. The Form 78, Form 60, and Form 40/45 contained 38, 34, and 30 items respectively. Munby et al. reported 29 items in the California F-scale and did not list any reliability estimates. The F-scale as appended in the Rokeach (1960) volume included just 29 items, eliminating the item "the true American way of life is disappearing so fast that force may be necessary to preserve it." Further, this item may have been considered as irrelevant to the Canadian sample. Robinson and Shaver (1973) reported reliabilities of the F-scale as follows:

> The reliability (split-half) on Form 78 over all groups was .74. Group reliabilities ranged from .56 to .88. Form 60 had a reliability of .87 over all groups tested with group reliabilities ranging from .81 to .91. Forms 45 and 40 had a reliability of .90 over all groups tested. Individual group means varied from .81 to .97 (p. 422).

The data in the Munby et al. report on the F-scale listed Michigan U. as one of the groups sampled. According to Rokeach (1960) it was Michigan State University.

The items from the several inventories were administered to "168 secondary school teachers of a variety of subject specializations." It seemed that the study was oriented to science teaching and science teachers. How this sample will assist for thar purpose is not clear.

Bivariate correlation coefficients were calculated between the total scores of the four scales and the random number variable. Conclusions based on the calculated statistics and testing of hypotheses seemed appropriate and well-founded. The investigators failed to relate these findings to those cited by Rokeach (1956) as cited in Robinson and Shaver (1973).

Intercorrelations among the D, E, and F scales with seven different samples, varying in size from 60 to 202 were cited in the Robinson and Shaver volume (1973). The D-E correlations ranged from .31 to .53, D-F from .54 to .77 and E-F correlations from .56 to .64. The results cited in the Munby et al. report are consistent with these findings.

While it was not important for this study, it would have been interesting to have had a brief description of the degree of dogmatism-authoritarianism in this sample of teachers.

The D scale was composed of 66 items with a range of possible scores of 66 to 462 (with the scoring conversion). The neutral position would correspond to a score of 264. Means above that would be interpreted as indicating a higher degree of dogmatism. The sample of 317 college students cited by Munby had a mean of 219, almost exactly the same as the Munby sample—220. Those values could be described as non-dogmatic, although not strongly so.

Similarly, the E scale scores could range from 40 to 280 with a neutral position corresponding to a score of 160. The student samples investigated by Rokeach had means ranging from 141 to 144, with the Munby teacher sample having a mean of 134. Again, both sets of data could be described as slightly non-enthnocentric.

Possible scores on the F scale range from 29 to 203 with a neutral position of 116. The data collected from students by Rokeach had mean values from 83 to 116. The Munby sample had a mean of 94—once again, similar to the Rokeach sample and below the middle position.

Performance Anxiety may be a somewhat different trait than Anxiety in general, although they initially seem to have considerable commonality. Rokeach (1960) did suggest that

> ...to the extent that a belief-disbelief system is closed, it represents a cognitive network of defenses against anxiety. This leads us to the simple hypothesis that those with relatively closed systems should manifest more anxiety than those with relatively open system (p. 347).

Rokeach obtained data relating to this hypothesis from a variety of groups in the United States and England. In all cases the "measure of anxiety used was 30 items from the Minnesota Multiphasic Personality Inventory (MMPI)." Some typical items were:

I work under a great deal of tension.
I have nightmares every few nights.
My sleep is fitful and disturbed.
I frequently notice my hand shake when I try to do something.

The correlations on the groups, with the number of subjects ranging from 60 to 207, between the Dogmatism and Anxiety scales ranged from .36 to .64, all very significant.

These results are very different from findings of Munby and colleagues between the Dogmatism scales (D, E, and F) and the PRCS, their measure of performance anxiety. A number of possible explanations of these differences may be possible, but are beyond the scope of this abstract.

The authors correctly identified this study as "an early portion of the effort to put an elaborated philosophical distinction to empirical test...." Comprehending philosophical considerations is a weighty problem by itself. Adding two methods of assessing traits which are difficult to conceptualize makes the task truly immense. The investigators have done an admirable job in this difficult domain. It is hoped that they continue to contribute to this research endeavor.

REFERENCES

Adorno, T. W.; E. Frenkel-Brunswik; D. J. Levinson; and R. N. Sanford. The Authoritarian Personality. New York: Harper & Row, 1950.

Campbell, D. T. and D. W. Fiske. "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix." Psychological Bulletin, 56:81-105, 1959.

Kimball, Merritt E. "Understanding the Nature of Science: A Comparison of Scientists and Science Teachers." Journal of Research in Science Teaching, 5:110-120, 1967-68.

Munby, A. H. "The Provision Made for Selected Intellectual Consequences by Science Teaching: Derivation and Application of an Analytical Scheme." Unpublished doctoral dissertation, University of Toronto, 1973.

Nagel, Ernest. The Structure of Science. New York: Harcourt, Brace and World, 1961.

Robinson, John P. and Phillip R. Shaver. Measures of Social Psychological Attitudes. Ann Arbor: Institute for Social Research, 1973.

Rokeach, Milton. The Open and Closed Mind. New York: Basic Books, 1960.

Rokeach, Milton. "Political and Religious Dogmatism: An Alternative to the Authoritarian Personality." Psychological Monographs, No. 425, 70(18), 1956.

EDITORS' NOTE: The complete report of the Munby et al. study is available in The Alberta Journal of Educational Research 24(2):69-80, June, 1978 (personal communication from Hugh Munby).

ATTITUDE STUDIES

22/23/24

Novick, Shimshon and Dina Duvdvani. "The Scientific Attitudes of Tenth-Grade Students in Israel, as Measured by the Scientific Attitude Inventory." School Science and Mathematics, 76(1): 9-14, 1976.

Descriptors--*Educational Research; International Education; Science Education; *Secondary School Science; Secondary Education; *Scientific Attitudes; *Surveys; Scientists

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

Purpose

In the recent past, the National Association for Research in Science Teaching has underscored the value of attitudinal research in science teaching by elevating it on the priority listing. This emphasis has produced a number of studies on the affective component of science teaching including the present study by Novick and Duvdvani who studied scientific attitudes in tenth-grade Israeli school students who had been exposed to the "big three" science courses; biology, chemistry, and physics.

Rationale

The authors presented no rationale for this study. The reviewer infers that the rationale was a comparison of scientific attitudes of students in public schools in two nations with somewhat similar economic orientations. Of all the prior research done in the area of scientific attitudes, only that of Moore, the original author of the Scientific Attitude Inventory (SAI) instrument, was presented.

Research Design and Procedure

The design of this study was a non-equivalent comparison group study using randomly chosen classes (n = 684) from within the strata of different school-types in Israel. The comparison group was drawn from literature reporting the results of a test with ninth-grade students in the United States and with a college sample instructed in a physical science course for non-scientists. The data on the comparison group were gathered in the United States during the 1970s while the data were gathered from the Israeli schools in 1973. The sample sizes of intact classes were essentially equivalent in the Israeli and comparison schools. The comparison and the Israeli schools were not sampled randomly.

The instrument used was the Scientific Attitude Inventory (SAI) which is a Likert-type instrument designed to measure the extent to which a student assumes six pairs of positive or negative scientific attitudes. The instrument contains 60 items, 6 per attitude. The first three attitudinal components bear an intellectual orientation while the remaining three attitudes reportedly bear an emotional attitude toward science. The reliability and validity of the original instrument are reported by reference only. The instrument was translated into the Hebrew language and reportedly validated in the same manner as the original instrument. No data were given on this validation except that the overall reliability coefficient from the Israeli study was 0.58.

The data collected included means and standard deviations which were computed and compared for the total Israeli sample and the individual United States samples. No statistical tests of means were performed. The summary means and standard deviations were reported for the total test, the individual subscales and the two subscales which were summed across the first three and the last three scales. No reliability or validity data were reported on any of the subscales.

Findings

The authors reported that the Israeli mean scores compared favorably with the mean scores for the U.S. data for the total and all subscales. The authors inferred that Israeli students consider scientific work interesting and rewarding to a greater extent than did the U.S. sample. They speculate this difference might be due to cultural-technological interactions which were a function of the 1970s in the United States and 1973 in Israel. They also speculate the heightened awareness in the United States of environmental effects as a potential contributor to this supposed discrepancy.

The authors also inferred that the need for public support and understanding of science was appreciated more in the United States than it was in Israel.

Interpretations

A portion of the findings and discussion addressed the unidimensionality of the attitudes being measured. The argument was put forth that a unidimensional attitude would result in small differences between mean scores for the positive and the negative statement of an attitude. In fact, the data showed the difference between the means on positive and negative statements for each attitude to be large. No statistical tests of significance were reported on these mean scores. As a further excursion into this issue of unidimensionality, correlations between positive and negative statements for each attitude were made. Three of the six correlations were significant at the .01 level and two of these were positive.

The authors questioned the unidimensionality of the instrument and referred to Moore's interpretation which centers on the notion that students do not seem to understand the nature of scientific explanations in the form of theories and laws regarding the first attitudinal component. This statement calls into mind the very close and not fully understood link between intellectual knowledge and attitudinal development. The authors concluded that the development of scientific attitudes required much more explicit attention in current Israeli science programs.

ABSTRACTOR'S ANALYSIS

As Sir Isaac Newton indicated, greatness in part stems from standing on the shoulders of giants who have gone before. Similarly those dealing with education in an empirical manner would be well advised to consult the research and theoretical literature as a basis for their work. For example, Wagner and Sherwood (1969), Festinger (1957), Shrigley, Riley and Johnson (1976) and Shrigley (1978) suggest four models of attitude development from a theoretical basis.

In the design, there is little description of similarities and differences between the comparative samples. This lack of information severely restricts the ability to make inferences and generalizations. For example, during 1970, tremendous change was occurring in the cultural and social structure of this country. Such changes included a growing awareness of the deleterious effects of science and technology on the environment, an emphasis on the use of technology for war time purposes, and a sharp drop in the demand for some scientifically trained persons, such as engineers. Although space is limited, the authors might have provided a brief statement of the nature of the samples that were used as a comparison.

Under the heading of instrumentation, the reviewer would like to comment on cultural validity, reliability, and unidimensionality. Although the instrument was translated verbally from English into the Hebrew, there was no report on a cultural translation. It is well documented that tests can carry biases when used in a culture for which they were not designed. Cultural bias can even occur within the same culture. Witness the trend in recent years towards technologically and practically based research versus theoretical studies in the sciences, a result largely of changes in funding patterns by the federal government. One can only speculate as to the cultural biases that may have been introduced in this particular study.

Although the authors indicated the use of similar procedures in validating the translated instrument, no data were presented to indicate the validity of the instrument or its subscales. The reliability of the total instrument was 0.58 which is clearly questionable. Standard errors of measurement were not presented in this report. Although conclusions were speculated regarding the subscale scores, neither the reliability nor the validity of any of the subscale scores was mentioned. Coupled with the fact that no statistical tests were performed with the exception of the correlation coefficients, it appears that any conclusions or generalizations are truly speculative at this point.

The authors raised the question of the unidimensionality of the attitude underlying this particular instrument. The reviewer, however, would like to raise the question of the unidimensionality of the attitude scales embodied in the SAI. Were the scales unidimensional, it would seem that mean scores on the positive statement and negative statement would be inversely correlated. However, the correlation between positive and negative pairs across all six subscales indicates little, if any, significant correlation and of the three significant correlations, only one is in the negative direction. The argument presented by the reviewer is in direct opposition to the arguments made by the authors which predict positive correlations on the scaled pairs.

The problem of unidimensionality is a severe one and is further supported by the instrument itself. Scale 4a, for example, starts out with the statement that science is an idea-generating activity while scale 4b, the supposedly negative view of the same concept, suggests that science is a technology-developing activity. Theoretically, if one rates statement 4a highly, he would then rate 4b low. The face value of these two statements, however, is such that the student with a strong understanding of public funding patterns would rate both these statements highly. The naive student who had memorized a definition of science versus technology might rate 4a high and 4b low.

Related to the issue of unidimensionality is absence of data on the uniqueness of subscales 1-6 or the two larger subscales on attitude versus intellect in the instrument itself.

The conclusions based upon the subscales are unwarranted in the light of the absence of reliability figures for subscales and the lack of tests of significance.

Assuming the above problems are straightened out, the reviewer would like to make three suggestions. First, these data are correlational and subject to the influence of other variables. Some of these can be measured and partialed out, others cannot. Those which can be measured and partialed out using partial correlation techniques should be. Somewhat more precise statements could then be made.

Second, an analysis of response by school types (vocational, academic, religious, etc.) would appear fruitful.

Discrepancies between paper and pencil responses and actual behaviors have been documented in other areas (e.g., marketing research). If these discrepancies occur in attitudes toward science, the inferences, conclusions, and resulting courses of action may be erroneous.

The knowledge component of attitude has long been questioned. If the first three subscales of the SAI do indeed measure an intellectual (i.e., knowledge acquisition) component and the next three are attitudinal, it may be possible to use the first three subscales to partial the intellectual component from the attitudinal, leaving a more pure assessment of attitude. Empirical support for this distinction, perhaps through factor analytic techniques, should be employed.

REFERENCES

Festinger, L. A Theory of Cognitive Dissonance. Stanford, Calif.: Stanford University Press, 1957.

Shrigley, R. L., J. Riley, and T. Johnson. "Generating Theoretical Bases for Attitude Research in Science Education." A paper presented at the meeting of the National Association for Research in Science Teaching, San Francisco, 1976.

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.

Wagner, R. W. and J. J. Sherwood. The Study of Attitude Change. Belmont, Calif.: Brooks-Cole Publishing, 1969.

Osborne, Roger. "Using Student Attitudes to Modify Instruction in Physics." Journal of Research in Science Teaching, 13(6): 525-531, 1976.

Descriptors--Attitude Tests; College Science; *Curriculum Evaluation; *Educational Research; Higher Education; *Laboratory Experiments; *Physics; Science Education; *Student Attitudes

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Marvin Bratt, The Ohio State University.

Purpose

The purpose of this study was to gather information concerning students' perceptions of laboratory work in a university-level physics course. This information could then be used as a basis for decision-making in the planning of future courses.

Rationale

The study as such was not related to any previous studies in the area. No contextual framework or model was mentioned. However, the author quotes Stufflebeam's definition for evaluation as "the process of delineating, obtaining and providing useful information for judging decision alternatives."

Research Design and Procedures

This was a formative evaluation study involving two measures of student attitude over a two-year period in order to modify a college-level physics course. Perhaps the studies could be described as "post-test only" designs. In the first study, conducted in 1973, 53 students enrolled in the course completed five units of work. Upon completion of the course, these subjects completed a 26-item semantic differential attitude inventory. These data were used to construct profiles which were compared to similar data collected the following year, 1974.

The 1974 study was conducted with 80 subjects, also students in the physics course. In short, the researcher collected attitude data with a semantic differential; he then changed or in some way modified three of the instructional units (laboratory work). The second year he gathered data using the same instrument, constructed profiles and looked for differences across the adjective pairs.

In addition to the above, the researcher also subjected the data to factor analysis. Using the Kaiser varimax method he rotated the data orthogonally and found four significant factors and one which was mixed. The factor analysis data and the profiles for each year were illustrated graphically.

Findings

The findings in this study may be summarized as follows:

1. 50 percent of the students surveyed upon completion of the physics course had changed their attitudes toward physics in a positive direction, 20 percent considered their attitudes more negative.

2. Laboratory studies appear to influence student interest in planning experiments, analyzing the results of experiments, appreciating the nature of physics and interest in physics.

3. Varying the activities in the laboratory setting significantly ($p \leq 0.01$) changed the mean values on experiment one bipolar adjectives: short-long, straight-forward theory-complex theory and varied-monotonous. These data are displayed in graphical form as profiles.

of research in the affective domain. Perhaps more precision is required in the area of measurement. There is, however, no question that the data collected can be useful in decision making by physics instructors in the university setting.

Several questions could be asked about the study as it stands. The information presented in tabular form concerning student opinion is vague. There are no defined parameters from which one could argue student opinion and the most effective learning environments (Table 1). The author spent valuable space describing the procedures used in the factor analysis of the instrument which does not relate to the objective of his study. It would seem that the factors he described are related to each other, therefore begging the use of an oblique solution in preference to the orthogonal solution chosen. Too many studies of this type fail to describe precisely what variables were changed in the instructional sequence. Even studies which use Harvard Project Physics, Chem-Study Chemistry or BSCS Biology as experimental treatments do not describe exactly what has happened so that future researchers could replicate the studies. In this study, changes were made in the first laboratory activity—it would seem that the first laboratory experience would not be the "best" unit of work to attack. One additional comment that is pertinent...perhaps a pre-post testing design would have been more informative in this study, especially if the researcher was committed to making major investments of time and energy involving changes in the course activities.

The study does add a dimension to the area of research on attitudes towards science and science teaching. It is a demonstration that information gained in research can be applied to discrete and pressing problems. The factor analysis data tend to support other studies in this area, especially when viewed along with profile data. The negative feelings demonstrated towards mathematical aspects of physics have been a serious concern to researchers for some time. Normally one looks for ten subjects/cell when dealing with factor analyses on data matrices. A 26 x 26 matrix would require roughly 6500-7000 subjects. Cattell (1973), however, has suggested that factor solutions may indeed be valid for much smaller samples or subject/cell ratios. Perhaps further research could also use cluster analysis programs or multidimensional scaling techniques as a comparison to factor analytical models. In essence, the procedure leads to the assumption of construct validity for the attitude inventory (Bratt, 1978).

Just as a point worth noting, the researcher did not describe the statistical tests used to determine significant positive changes in attitudes—this would have been meaningful. Overall, the report is very interesting. Some of the details are buried and somewhat confusing. Perhaps more discussion of exactly what changed between 1973 and 1974 would be more helpful to the reader.

REFERENCES

Bratt, Marvin and Alfred DeVito. "The Development and Field Testing of an Inventory of Humanistic Attitudes Towards Science Teaching." Journal of Research in Science Teaching, 15(6):551-557, 1978.

Cattell, R. B. Factor Analysis. Westport, Connecticut: Greenwood Press, 1973.

Herron, J. D. et al. "A Summary of Research in Science Education, 1974." Science Education Supplement, 1976.

Mallinson, George. "A Summary of Research in Science Education, 1975." Science Education Supplement, 1977

Renner, J. W. et al. "A Summary of Research in Science Education, 1976." Science Education Supplement, 1978.

Shrigley, R. L. "The Function of Professional Reinforcement in Supporting A More Positive Attitude of Elementary Teachers Toward Science." Journal of Research in Science Teaching, 14(4): 317-322, 1977.

Descriptors--*Attitudes; Change Strategies; Educational Research; Elementary Education; *Elementary School Science; *Elementary School Teachers; Inservice Teacher Education; *Science Course Improvement Project; Science Education; *Surveys

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Chester E. Raun, Temple University.

Purpose

The purpose of this study was to:

1) Survey the attitudes of elementary school teachers toward the teaching of science.

2) Classify the responses with 8 or 10 broad barriers to science teaching.

3) Examine the barriers for a possible relationship to one or more of the four functions of Katz's (1960) functional approach to attitude change.

4) Recommend whether or not Katz's approach seems to be a valid theory on which hypotheses could be established and experimental studies designed to test attitude change of teachers.

Rationale

The investigator attempted to construct a relationship between his observations that "there seems to be little professional reward for the teacher who teaches science to children" and that to be successful at the implementation of such science programs as SAPA, SCIS or locally developed curriculum, there might be a need to analyze "other" professional needs, specifically those that spring from an attitudinal source. He raised the following questions:

1) Could the lack of professional reinforcement be the source of the less than positive attitude elementary teachers have toward the teaching of science?

2) Is there a theoretical basis in the attitude literature of the social psychologist that seems to support the hypothesis that teachers may need greater professional reward for science teaching?

3) Could Katz's functional approach to attitude change be a valid theoretical basis for the investigation of attitudes in teacher education?

The investigator identified a basic assumption of the functional approach as drawn from Wagner and Sherwood (1969). This assumption was "that attitudes develop and change as they serve to promote or support goals of the individual; that is, attitudes are instrumental to the person's satisfaction of his needs."

The four functions of Katz's functional approach are: (1) utilitarian, (2) knowledge, (3) value-expressing, and (4) ego-defensive. The investigator gave a description of each which appears to be his interpretations of Katz's approach as applied to teachers. For example, the investigator stated that the utilitarian function of Katz's approach suggests that teachers will modify their attitude toward the teaching of science when the change maximizes rewards and minimizes punishment. He suggested that perhaps teachers are not as reinforced by the school community for teaching science as they are for teaching reading, math and other subjects. Or, administrative restraints may serve indirectly as punishment to the teacher who teaches science. Therefore, changing teachers' attitudes could require external rewards in the form of materials, equipment, or smaller classes.

Similar logic was used to describe the other three functions with the investigator acknowledging that the ego-defensive function may be a limitation of the study.

There is no attempt to cite previous research related to the investigator's study. There were however, a number of related studies completed during the 1950s and the 1960s including studies by this abstractor and colleagues.

Research Design and Procedure

The investigator developed a simple, open-response statement to survey the attitudes of teachers concerning barriers to science teaching. The questionnaire—instrument was piloted with 239 inservice teachers. The instrument included:

1) The statement: "I will teach more science if/when..."
2) The directions: "Complete the statement in as few words as possible." "Do not identify yourself."
3) A question as to grade level taught.
4) A question as to whether the teacher has used SAPA, SCIS, or ESS materials.

To establish interrater reliability, the investigator randomly selected 50 teacher statements from the total responses and presented these to a jury of four teacher educators who were to use a procedure identical to that of the investigator, namely:

1) Classify each of the 50 responses in one of the following ten categories: software, hardware, time, teaching skills, administrative restraints, other subjects, assistance, student, miscellaneous, or I already teach science.
2) If a response is really a multiple response with more than one barrier (category) noted, record only the first barrier.
3) When the response is "materials," assume that the respondent means "software" and not "hardware."

A mean percentage agreement among the jurors of 82.5 was reported.

Consistency of results in ranking of the categories led the investigator to conclude that the instrument was reliable. In the pilot sample 85 percent cited a barrier (category) with the rankings as identified in the first procedure above.

Assuming that a change in the free response open-ended statement from "teach more science" to "teaching science more effectively" would raise the percentage of respondents citing barriers, the instrument was administered to a survey sample of 449 teachers representing a cross section of elementary and special education. A higher percentage of teachers were from grades 4-6 than K-3 with at least 20 percent using SAPA, SCIS, or ESS materials in their classrooms.

The sample represented teachers from 14 school districts (7 rural, 5 suburban, 2 urban) and 13 inservice classes at 7 colleges and universities in eight states.

The data from both the pilot study and the survey sample were recorded in a summary table indicating a percentage of teacher responses for each of the ten categories and a relationship to one of Katz's functions.

Findings

The investigator classified all of the 449 teacher responses and judged which of Katz's four functions seemed most congruent with each of the categories. The results were:

1) Software, hardware, time, removal of administrative restraints, and assistance accounted for 78 percent of the teachers' responses and seemed to be supported by Katz's utilitarian function.

2) Teaching skills, which seemed to be supported by Katz's knowledge function, accounted for 10 percent of the responses.

3) Placing science on a level below other subjects and teachers' questioning of the child's need for or interest in science sccounted for four percent of the responses and seemed to be supported by Katz's value-expression function.

4) There were no responses that seemed to be supported by the ego-defensive function.

Interpretations

The investigator concluded that:

1) Most of the teachers' responses seemed to be supported by one of Katz's three functions: utilitarian, knowledge, or value-expression.

2) The knowledge and value-expression components functioned at a low level.

3) The survey failed to reveal the ego-defensive function.

In addition to these conclusions the investigator extracted several implications regarding the inservice teacher:

1) That the attitude of elementary school teachers, and therefore the improvement of the teaching of science, rests primarily in forces outside the control of teachers themselves.

2) That principals and supervisors should reinforce a positive attitude by exercising leadership in integrating science materials into the local science teaching milieu by providing accessibility of materials and by planning to replace expendables methodically.

3) That school officials and science educators could team up to provide teachers with such teaching skills as oral questioning, teaching strategies, and a classroom climate amenable to the teaching of investigative science.

4) Aides and/or administrative help is needed to assist teachers in the preparation of science materials.

5) There is a need to establish science as a legitimate subject in the local school curriculum.

The investigator indicated that the study did not reinforce traditional views of teacher's being altruistic nor that professional motivations come from within. The study did not reveal personal dislike or discomfort associated with teaching science and the investigator acknowledged the existence of other undisclosed motives that interfere with some elementary teachers embracing science.

ABSTRACTOR'S ANALYSIS

The investigator attempted to explore both the question of professional reinforcement as a means of creating a more positive attitude on the part of elementary teachers toward science and the possibility of a theoretical basis for investigating attitudes in teacher education.

In developing his rationale the investigator did not draw upon preceding research which studied elementary teacher attitudes toward science. Several such studies exist and their strengths or weaknesses may have aided in the development of this study. For example, in the knowledge function the investigator includes teaching skills, knowledge, method, experience, and interest. A study by White, Butts and Raun (1969) indicated that previous teaching experience does not appear to be related to competence in science but does appear to be related to a teacher's attitude. In another study, Butts and Raun (1970) found that few or no formal science courses for elementary teachers resulted in a more positive attitude towards science. The findings of these and other studies are certainly related to this study.

The investigator in discussing the value-expressive function suggests that sex may make a difference since "85 percent of elementary school faculty personnel are females who view science as a male enterprise." This is entirely possible, yet in the study, both pilot and survey, there is no indication that the sex of the respondents was obtained. If the sample was entirely female, are the results biased?

Not only was omitting the ego-defensive function an unnecessary limitation of the study but a question concerning a teacher's dislike or fear of science should have been included. There should also have been an attempt to collect factual information beyond grade level taught and use of SAPA, SCIS, or ESS materials. A dislike or fear of science may be due to many variables, not the least of which is a lack of science content knowledge. Questions to seek information on content hours in science, type of science instruction received, etc., would have been appropriate.

The open question survey instrument is helpful if depth information is sought or there is uncertainty as to what the responses might be. It requires a large time investment in interpreting and categorizing responses. It is also inherently subjective and, as the investigator acknowledges, he classified responses into ten broad categories and he judged which of the four components of Katz's functional approach supported the categories. The consistency of rankings of the categories was used by the investigator to justify reliability of the instrument. Consistency is normally an indicator of reliability but Hyman (1968) considers the open question survey instrument as having low reliability because of inherent subjectivity.

In considering the category of "I already teach enough science," which was used in the pilot sample, the investigator reported a response rate of 15 percent which left 85 percent selecting one of the barrier categories (softward, hardware, etc.). Desiring to raise the percentage selecting a barrier category the base question was changed from "I will teach more science if/when..." in the pilot run to "I will teach science more effectively if/when..." in the survey. This resulted in a category which was now labeled "I already teach science effectively; am already satisfied" with a response rate of 5 percent. The investigator accomplished his purpose but left an interesting question. Was it possible that some of the respondents of the pilot group, who indicated that they already taught enough science, had indeed taught none which to them was enough? One may ask how do we know that all of the respondents in both groups taught science?.

The investigator attempted to establish interrater reliability of the category classification process. In doing so the raters were told to use the investigator's procedures including the statement that any response which said "materials" was to be assumed to mean "software" and not "hardware." Not only is it highly questionable to make such an assumption but it has been this abstractor's field experience that teachers who were prepared for and use any of the newer programs of elementary school science, such as SAPA, SCIS, or ESS, will insist on "hardware" when asked what they want in the way of materials. The assumption of the investigator becomes more acute when he indicates that 20 percent of the survey sample or approximately 100 teachers were using SAPA, SCIS, or ESS materials in their classrooms. How can one be assured that the response percentages for the software category (38 percent) and the response percentages for the hardware category (18 percent) are truly reflective of the "intent" of the teachers' responses? How many of this 20 percent of the sample had inservice preparation to implement and teach the identified programs?

The investigator does not indicate how many questionnaires were mailed to the 14 school districts nor the number returned from all 14 school districts. There is no response ratio given and we have no clues as to the characteristics of the respondents versus the non-respondents. Random sampling was not used nor was there any indication that the sample was equalized by random assignment.

The conclusions, stated in three brief sentences, provide a very broad response to the first three stated purposes of the study but do not address the fourth purpose. There is no indication that the investigator views Katz's approach as a valid theory for establishing hypotheses and designing experimental studies to test attitude change of teachers.

On a purely philosophical basis this abstractor could concur with most of what the investigator identifies as implications of the study. However, to draw these implications from the data of the study is not supported by the abstractor. For example, the investigator states, as an implication, that "respondents in this study suggested the use of aides for this purpose," —the purpose being time to prepare and assemble science equipment. A reported two percent of the responses made this suggestion which was hardly a significant number and a questionable percentage for the stated implication. In another instance the investigator indicated that "although the percentage was not high, respondents openly sanctioned the need to establish science as a legitimate subject in the local school curriculum." There are no data or response categories provided in the report to support this statement. Similar comments may be made regarding other stated implications. By and large, the investigator has gone far beyond the data provided.

In the opinion of the abstractor, this study could have been strengthened significantly. Aside from previous comments it may be suggested that the open question format of the pilot survey could have provided the basis for a closed question format. Obtaining factual information about the respondents could have been combined with survey data and analyzed by correlational techniques. Such analysis might have provided more definitive clues as to why certain categories received a larger percentage of responses and which categories and related factual information are functionally indicative of teacher attitude.

REFERENCES

Butts, D. P. and C. E. Raun. "A Study of Teacher Attitude Change," Chapter 10. Research and Curriculum Development in Science Education, D. P. Butts (ed.). Austin, Texas: University of Texas publication number 7023, December, 1970.

Hyman, J. L., Jr. Research in Education. Columbus, Ohio: Charles Merrill Publishing Co., 1968.

Katz, D. "The Functional Approach to the Study of Attitudes." Public Opinion Quarterly, 24:163-204, 1960.

Wagner, R. and J. Sherwood. The Study of Attitude Change. Belmont, Calif.: Brooks/Cole Publishing Co., 1969.

White, M. A., C. E. Raun and D. P. Butts. "A Study of Contrasting Patterns of Inservice Education." Science Education 53(1):13-20, 1969.

ACHIEVEMENT

45/46

Bennett, Roy Marshall. "Effects of Previous High School Programs on Achievement in College Biology." Journal of College Science Teaching, 4(4):242-244, 1975.

Descriptors--*Academic Achievement; *Biology; *College Science; *Educational Research; Higher Education; Science Education; *Secondary Education; Secondary School Science

Expanded Abstract and Analysis Prepared Especially for I.S.E. by N. Jean Enochs, Michigan State University.

Purpose

The stated purpose of this study was to determine the influence of particular high school biology programs and several background variables on achievement in beginning college biology.

Rationale

On a subjective basis, most college biology teachers feel that BSCS biology programs at the high school level provide a better preparation for beginning college biology than do other biology programs. A study of achievement in college biology by students with different high school background is needed to provide evidence on which to base a conclusion.

Design and Procedure

The subjects used in this study were 857 freshmen students enrolled in beginning biology during one quarter at Iowa State University. The students were distinguished on the basis of five factors: (1) which of six high school biology programs they had experienced (BSCS Blue Version [BV]; Yellow Version [YV]; and Green Version [GV]; Holt, Rinehart and Winston's Brown Version [Brown]; other high school biology programs [OB]; and no high school biology [NB], (2) the college in which they were enrolled at the university, (3) the type of college laboratory in which they were enrolled, (4) whether or not they had taken high school chemistry, and (5) sex. Achievement of different groups of students in beginning college biology was measured by a final grade which was based on two tests and a final exam.

The manipulated variables of the study were the five student description factors; the responding variable was achievement in beginning college biology. It appeared that an attempt was made to adjust for a scholastic ability and achievement variable though these variables were not mentioned.

Initially, product-moment correlations were calculated for each high school biology program and raw final grade. An exploratory multiple regression analysis including the following variables appears to have followed: final grade in college biology; rank in high school; scores on the Minnesota Scholastic Aptitude Test (MSAT), English Placement Test (ENG). Mathematics Placement Test (MATH): and the five student description factors listed previously. The MSAT and MATH were found to be significantly related to final grade, and two significant first order interactions were identified.

The primary analysis was an analysis of covariance which used MSAT and MATH scores as covariants with final grade to produce adjusted mean final grades. All significant differences between levels of the main effects and the two interactions were presented. Finally, a multiple comparison investigation of all possible combinations of high school biology programs was performed using least significant differences (LSD) tests.

Findings

The raw mean final grades for each of the high school biology groups were: BV = 2.52; YV = 2.46; NB = 2.31; Brown = 2.30; GV = 2.29; and OB = 2.26.

The analysis of covariance indicated significant (0.05 level of confidence) differences between the final grade means of the six high school biology groups and between a high school biology group by laboratory interaction. Significance at the 0.01 level is reported for differences of the mean final grade for the variables: chemistry background, sex, and a chemistry by sex interaction. Data for the interactions are shown in Table 1; no other main effects or interactions were significant.

The LSD multiple comparison of final grade and biology group showed significance at the 0.05 level for GV vs Brown and OB vs NB; 0.01 level significance was reported for BV vs OB, GV vs OB, YV vs Brown, YV vs OB, and Brown vs OB. Mean grades for all other combinations showed no significant difference.

Interpretations

The stated conclusions were that "students who had participated in the OB versions achieved significantly better," with the YV next, and GV third. No differences were found "between colleges, or between students who took a laboratory, based upon high school background. High school chemistry did produce highly significant results..., Males differed from females at the one percent level compared to females." About the two significant interactions, it is concluded that, "laboratory for majors seemed to be a contributing factor" and "males who had high school chemistry were significantly different." Finally, MSAT and MATH were found to be better predictors in college biology than rank in high school.

TABLE 1

MEAN FINAL GRADES FOR BIOLOGY BY LABORATORY INTERACTION AND CHEMISTRY BY SEX INTERACTION*

Group	No Lab		Major's Lab		Non-Major's Lab	
	Raw	Adjusted	Raw	Adjusted	Raw	Adjusted
BV	-	2.63	-	2.72	-	1.90
GV	-	2.63	-	3.23	-	1.96
YV	-	2.16	-	3.06	-	2.05
Brown	-	2.20	-	2.50	-	1.79
OB	-	2.74	-	2.69	-	2.30
NB	-	1.71	-	2.20	-	1.83
Combined Groups	2.16	2.13	2.50	2.63	2.32	2.25

	Male		Female		Combined Sex	
	Raw	Adjusted	Raw	Adjusted	Raw	Adjusted
Chemistry	-	2.40	-	2.43	2.41	2.53
No Chemistry	-	2.47	-	2.05	1.94	2.15
Combined Chemistry	-	2.55	-	2.13	-	-

*This table was prepared by the abstractor.

ABSTRACTOR'S ANALYSIS

Evaluating the effects of high school curricula on achievement in college science is a task worthy of science education research, but it has proved to be a difficult one. Previous studies of "new" curricula (e.g. PSSC, CHEM Study, BSCS) and "traditional" curricula have utilized several procedures and yielded inconsistent conclusions as to the effect of these programs on achievement in college science. Most studies report no significant differences in achievement in college, based on high school science curricula (Bajah, 1972; Garrett, 1968; Hendricks et al., 1963; Hudik, 1969; Mott, 1970; Ogden, 1976), while a very few claim such significance (Cothingham, 1970). The omission of related literature and an historical context limits the significance of this study to that of an initial investigation.

Although a general design and series of analyses are suggested above, the procedures, analyses and findings of this study are not clearly presented or interpreted in the paper. The study would have been greatly improved by more specific, initial planning. Initial analysis using raw final grades indicated that students who took high school chemistry had a mean grade of 2.41 while those who did not averaged 1.94 (Table 1). Although this result was said to be "expected," no mention was made of the possibility and importance of controlling for the distribution of these two groups throughout the biology groups. The same analysis indicated raw mean grades for the different college laboratory groups as follows: no lab - 2.16; major's lab - 2.50; non-major's lab - 2.32 (Table 1). These were shown to be significant at the 0.05 level. Also presented in the table is the result of a later analysis of covariance which showed that sex had a significant effect on the mean grade, males averaging 2.55 and females 2.13 (although inspection of the breakdown figures above this score in the table suggest that an error has been made; other data in the paper verify that the effect of sex on final grade was significant at the 0.01 level.) All three of these variables were known to have a significant effect on final grade, but no mention was made of techniques used to adjust the mean grade on this basis, nor was an effect on the validity of the results acknowledged. Uncontrolled distribution of these variables over the biology groups destroys the validity of the statistical procedures used.

Achievement in college biology as measured by raw mean grades for the six biology groups was profoundly effected by the adjustment of the means based on MSAT and MATH scores, e.g. the OB group which had the lowest achievement based on raw means was said to have the highest achievement after the means were adjusted. Despite the magnitude of this data manipulation, no rationale, no description of how the multiple regression analysis was carried out and no data resulting from the regression analysis were presented. The actual adjusted mean grades for each biology group are not reported; they are said to be presented in a specific table, but they do not appear. Without these means only difference scores are available to measure the affect of the different variables on final grade; the directions of the differences found by the analysis of covariance and the LSD multiple comparison cannot be inferred from data presented.

The generation of the all-important responding variable, mean grade in college biology, was not clearly described. The relationships of the different laboratories to the biology course and their possible direct effect on the final grade was not discussed nor was it clear that all the students were enrolled in the same beginning biology course. Additional ambiguity was introduced by a few apparently typographical and technical errors such as the statement that "the better student took YV" when the data being interpreted specified that the BV group had the higher mean grade. While the style of writing imposed by many journals due to space limitations often requires omissions and cryptic phraseology, it is important to include information which will allow a reader to reconstruct the analysis and, if possible, to carry out other analyses of the data that have been obtained.

This paper deals with an important topic and contains valuable data, but the data analysis is inadequate to support the conclusions presented. The literature describing investigations of this type suggests several approaches to the analysis of data; I recommend, as a model, the paper by Hendricks (1963). Several additional student variables have been analyzed for effect; age, mathematics background, and SAT scores as well as chemistry and sex have been reported to be significant (Ogden, 1976). I would be very interested in seeing the data reported in this paper reanalyzed to remove the difficulties cited above. It would make a valuable contribution to science education and, more specifically, to high school textbook selection committees.

52

REFERENCES

Bajah, Samuel Tunde. "A Study of the Relationships Between Selected Variables in High School Science Education and Academic Achievement in the Freshman Chemistry Program at the University of South Dakota." University Microfilms, Ann Arbor, 1972.

Garrett, Gordon Ronald. "Comparative Effects of Various High School Biology Course-Content Backgrounds on Achievement in College Biology." University Microfilms, Ann Arbor, 1968.

Hendricks, B. O'Neal; Charles L. Koelsche; and Joseph C. Bledsoe. "Selected High School Courses as Related to First Quarter Chemistry Marks." Journal of Research in Science Teaching, 1(1): 81-84, 1963.

Hudek, Albert D. "The Relative Effects of PSSC Physics and Traditional Physics on Achievement in College Physics." University Microfilms, Ann Arbor, 1969.

Mott, Roger Clyde. "An Investigation of the Relationship Between Students' Participation in Traditional or New Science and Subsequent Participation Performance and Perception of the Study of Science." University Microfilms, Ann Arbor, 1970.

Ogden, William R. "The Effect of High School Chemistry Upon Achievement in College Chemistry: A Summary." School Science and Mathematics, 76(2):122-126, 1976.

Cottingham, Charles Louis. "A Comparative Study of CHEM Study and Traditional High School Chemistry in Relation to Students' Success in College Chemistry." University Microfilms, Ann Arbor, 1970.

53

Hedges, Larry V. and Kenneth Majer. "Prerequisite Courses as Predictors of Achievement in the Natural Sciences." Office of Academic Support and Instructional Services, University of California, San Diego-La Jolla, California, February, 1976.
Descriptors--*Academic Achievement; *College Science; *Educational Research; Higher Education; *Natural Sciences; Prediction; *Predictor Variables; Science Education

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Herbert A. Smith, Colorado State University.

Purpose

The study investigated the validity of prerequisite lower division course grades as predictors of grades in upper division science major areas.

Rationale

As indicated in the study, prerequisite courses are normally required as a condition of enrollment in upper division courses. These requirements are usually justified on the basis of providing necessary content, providing practice in the use of skills in quantitative problem solving and as an efficient screening mechanism of students for advanced study. Is this a reasonable and valid practice? This study addresses itself to this question.

Research Design and Procedure

One hundred ninety-five college students who graduated from the University of California, San Diego, in the spring of 1975 were the subjects of this study. The students were in seven different majors: applied physics and information science, applied mechanics and engineering sciences, biology (two programs), chemistry, mathematics and physics. The procedure used in the study involved analysis of the students' records to obtain an average grade in the several prerequisite courses and the average grade in upper division courses in the students' major area. A correlation analysis was then made in which bivariant and multiple regression coefficients were obtained to provide a measure of relationship between the variables. The results are succinctly summarized in tabular form.

Findings

The findings clearly demonstrate that the prerequisite courses have relatively high validity as predictors of success in later upper division courses in the curricula included in this study.

Interpretations

The authors conclude that the prerequisites are justified on the basis of their findings. They suggest the usefulness of regression models to counselors and academic advisors in helping students make decisions about their future areas of study and, by inference, their probable success in pursuing more advanced study.

ABSTRACTOR'S ANALYSIS

This study is perhaps not highly original or innovative. But it is straight-forward, logical, well written and something of a model for clarity. Considering the small samples for a correlational study, it is surprising that all correlations were significant; most of them, including all the multiple correlation coefficients, at the one percent level. The study would appear to validate convincingly the use of prerequisites as predictors of success in more advanced courses in scientific fields. The fact that physics grades were the best predictors of later performance in four of the seven major areas (but strangely enough, not in physics) is intriguing. As indicated in the data, undergraduate mathematics grades appear to be better predictors for upper division physics grades, while lower division physics grades are better predictors of upper division mathematics grades. Presumably the differences between the coefficients are not statistically significant. It does, however, suggest that some additional study is warranted.

The question of the validity of prerequisites in other fields remains unanswered. On *a priori* grounds one would expect the validity of prerequisite courses as predictors of later success in upper division courses to be greater in fields with strongly developed structure such as mathematics, science and possibly economics or in fields with a relatively fixed sequence such as foreign languages, than in humanities, history or the less quantitatively oriented social sciences.

With the pressure to admit students of varied backgrounds, ages and experience, there is often pressure to waive or eliminate prerequisite courses.

One interesting question suggested by the study is the relationship of institutional policy on prerequisites and the quality of the institutional graduates. When prerequisites are either discouraged or easily waived, it would appear to be difficult to maintain standards in upper division courses. An extrapolation of this study would support this view, at least in science fields, and raises a general question about other academic and professional areas.

Pedersen, Arne A. and Judith E. Jacobs. "The Effect of Grade Level on Achievement in Biology." Journal of Research in Science Teaching, 13(3):237-241, 1976.

Descriptors--*Academic Achievement; Age Differences; *Biology; *Educational Research; *Instructional Program Divisions; Science Education; Secondary Education; *Secondary School Science

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Ellen Stephanie Simmons, Teachers College, Columbia University.

Purpose

According to Pedersen and Jacobs, the purpose of this study was "...to determine whether there was a difference in achievement between ninth- and tenth-grade students upon completion of a year of biology."

Rationale

After reviewing the literature on this topic, the authors concluded that the previous studies were conflicting in nature and narrow in scope. Hence, an investigation involving a large number of ninth- and tenth-grade subjects with diverse abilities was devised which controlled for many extraneous variables inherent in the earlier research.

Research Design and Procedure

This study was conducted over the course of one academic year involving all ninth-grade (684) and all tenth-grade students (721) from Tottenville High School, Staten Island, New York. An even mixture of ninth- and tenth-grade students was employed within each class. Furthermore, the classes contained pupils from a wide continuum of abilities; this larger spectrum of abilities was deemed necessary in order to offset those weaknesses found in the earlier research which tended to select only a small number of subjects and usually focused especially on the brighter students. Eighteen teachers were involved in the project; thus, extraneous variables, including teacher competency, classroom conditions, and course time, were controlled because both groups were exposed on a random basis.

At the completion of the year's study, all subjects were tested at the same time with absenteeism being minimal (16 ninth-grade, 20 tenth-grade). A comprehensive multiple-choice test covering the year's course of study was prepared by a committee of experienced biology teachers. While the test's reliability was not calculated, its validity was verified by the faculty involved.

The scores attained on this test were divided into the two class groups, and the mean for each group, as well as the range and the standard deviation, was calculated. A t-test for independent means with 0.05 level of significance was performed. After grouping the individual scores into intervals (selection was based upon teacher interests), a chi-square test was employed to determine if a significant difference existed at the 0.05 level between the grade distribution for ninth-graders and the grade distribution for tenth-graders.

Findings

No significant difference was found to exist between the mean scores for each grade level, and the difference in grade distribution was negligible. The t-test for determining the difference in mean scores showed a t-value of 1.72 which was not significant at the 0.05 level. The chi-square analysis for grade distribution resulted in 9.69 which was not significant at the 0.05 level.

Interpretations

Since no significant difference was found in either statistical analysis, the researchers inferred that ninth-grade students do have the background and the capacity to succeed in biology. Furthermore, the authors believe that their study supports those who have proposed that biology should be taught on the ninth-grade level; this conclusion is based on the authors' use of a larger number of subjects with diverse abilities.

ABSTRACTOR'S ANALYSIS

The authors of this research report convey their information succinctly, thoroughly, and precisely. The format used helps the reader clearly and quickly understand the totality of the investigation, the problem analyzed, the process used, and the results achieved. As a result of these many attributes, this investigation as reported has demonstrated that ninth-grade students should be permitted the opportunity to study biology.

This conclusion was reached after a careful study of the literature and after developing a research design devoid of the deficiencies in previous studies. Their investigation is unique in the employment of a larger number of subjects (using all ninth- and tenth-grade students with diverse abilities attending a large urban high school). Unfortunately, the authors did not quantify the "varying abilities."

However, several attributes are found in this study:

1. The mixture of ninth and tenth-grade students within individual classes—this controlled for subject matter presentation bias.

2. The final examination was prepared by a committee of biology teachers—this insured that the test content dealt with the entire year's curriculum and eliminated tester bias.

3. The final examination was administered simultaneously to all subjects—this alleviated any bias in testing conditions.

In addition to these attributes, careful attention was also given to the research design. The use of a t-test for evaluating and comparing the mean scores between grade levels was most appropriate as was the use of the chi-square for evaluating the significant difference in the grade distribution for the two groups. According to the authors, they chose unequal intervals with logical care, keeping in mind the interests of classroom teachers. While the chi-square test using unequal intervals was shown to be not statistically significant, it would be interesting to reanalyze the data using equal intervals except for the first from zero to 50. Since such great care was taken with the statistical analysis of the data, it seems unfortunate, however, that no quantitative techniques were employed to assess the reliability and validity of the achievement test (content validity was only subjectively determined by the biology faculty).

Therefore, the conclusion reached by the authors appears to be well-founded—that ninth-grade students would seem to have the background and the capacity to succeed in biology. The authors also suggest that as a result of being able to take biology in the ninth-grade, students will have more opportunities to select courses which best meet their individual needs and interests.

RESPONSE TO ANALYSIS

61/62

A RESPONSE TO THE ANALYSIS OF

Willson, V. L. and A. M. Garibaldi. "The Association Between Teacher Participation in NSF Institutes and Student Achievement," by E. J. Davis. Investigations in Science Education, 3(1):58-61, 1977.

by

Victor L. Willson
University of South Dakota

Volume 3, Issue 1 contained an abstract and analysis of an article first published in the Journal of Research in Science Teaching (1976). The abstract and analysis were prepared by Edward J. Davis. There are several distinct points of the criticisms of Davis which will be addressed in this response.

The data analysis reported was part of a five-year evaluation grant to Dr. Wayne Welch to study effects of selected NSF projects (Welch and Gullickson, 1973). The evaluation included an initial assessment of several hundred teachers in science and mathematics and concomitant assessment of one of their classrooms. Davis questioned the randomness of the sampling procedure in the selection of teachers, who were supposedly selected randomly by principals from a stratified systematic sample of schools in five geographic regions. No direct investigation of randomness was made, but indirect evidence was collected by examining the characteristics of the teacher sample. Number of years in teaching, number of credit hours of science and math, highest degree earned were compared between the ERIC sample (Schlessinger, F. R., et al., 1971) and the Welch sample. The two samples were quite similar (Dr. Wayne Welch, Personal Communication, 1978).

Another criticism of the article was centered on the construction of the achievement tests used. In mathematics the National Longitudinal Study of Mathematical Abilities (NLSMA) item set formed the basis for test construction. At grades eight and eleven, items were content categorized into number system, algebra, and geometry. Each was subdivided into four taxonomic levels: computation, comprehension, application, and analysis. Mathematics educators formed an expert panel to decide percentage contribution of each content subarea to reflect current curricula. Items were selected to have difficulties between .5 and .9 with point-biserial correlations greater than .3 with the original NLSMA scales. Further revision and a pilot test in public schools followed, with final revision. In science a comparable process was followed using National Assessment of Educational Progress (NAEP) Science Assessment items released from the 1968 assessment. Attitude and process measures administered in the study were discussed in other published articles and are not directly relevant to the paper. Technical descriptions of the tests are available from Dr. Wayne Welch, University of Minnesota, Minneapolis, MN 55455. See also Lawrenz (1972) and Sandman (1972).

The last point of criticism was considered most important by the reviewer. He suggested the gain associated with institute attendance was too little (about 2 points on the 40-item tests) to warrant significant attention or to recommend institute attendance. There are several features of the gain ignored by the reviewer. First, it is the average gain over a large number of classrooms. There have been few educational effects shown to be this large in any studies of achievement. Next, the gain can reasonably be compared with gains of similar magnitude on standardized science or mathematics achievement tests. Two points on a 40-item test represents a 5 percent absolute gain (and 10 percent on the base level of about 20 correct for the poorest groups). In Table 1 are listed comparable gains and their interpretations in terms of group means as published by the first two tests I had immediately at hand.

The magnitudes of gains for the classes in the Willson and Garibaldi study are comparable to very large average school gains on the STEP and ITBS tests. While grade equivalents are crude at best, the ITBS gains of about four months in grade 8 math are probably quite comparable to the gains of the STEP tests, which don't report grade equivalents. Any program which can produce an average four-month gain in classes is worthy of consideration, as was discussed in Willson and

Table 1: Comparable Mean Gains for NSF Data Set and Several Standardized Test Norm Groups

		Initial Mean Difficulty of Test for Comparison	Raw Gain	Percent Gain of Total Items	Grade Equivalent Gain	Percentile Rank Gain for School Means	Comments
1. NSF Data Sample (Willson and Garibaldi)							
Science (40 items)	Gr. 11	.56	2.14	5.35	--	--	Initial mean difficulty is the mean score for NO NSF institute attendance divided by number of items in the test.
	Gr. 8	.56	.95	2.38	--	--	
Math (40 items)	Gr. 11	.57	2.01	5.03	--	--	
	Gr. 8	.51	1.77	4.43	--	--	
2. STEP - Series II School Mean Norms (1971)							
Science (75 items)	Gr. 11	.55	4.01	5.35	--	+24%	Initial mean difficulty is the mean school raw score divided by number of items on the test as reported in test manuals.
(50 items)	Gr. 8	.63	1.19	2.38	--	+11%	
Math Basic Concepts (50 items)	Gr. 11	.49	2.52	5.03	--	+23%	
(50 items)	Gr. 8	.59	2.22	4.43	--	+21%	
Math Computation (60 items)	Gr. 11	.57	3.02	5.35	--	+31%	
(60 items)	Gr. 8	.62	2.66	4.43	--	+13%	
2. ITBS - Form 5 (1974) School Mean Norms							
Mathematics Concepts (30 items)	Gr. 8	.75	1.33	4.43	.4	+20%	
Math Problem Solving (30 items)	Gr. 8	.47	1.33	4.43	.4	+23%	

Garibaldi. The conclusion remains the same: participation in in-service institutes by teachers appears quite beneficial to student achievement.

REFERENCES

Iowa Tests of Basic Skills: Teacher's Guide for Administration, Interpretation and Use. Boston: Houghton Mifflin Co., 1971.

Lawrenz, F. "Test of Achievement in Science: Description and Copies, Form I and Form II Instruments." Project Report No. 2, Minnesota Research and Evaluation Project, University of Minnesota, March, 1972.

Sandman, R. S. "Development of the Mathematics Achievement Tests and Copies of the MAT, Form I and Form II Instruments." Project Report No. 4, Minnesota Research and Evaluation Project, University of Minnesota, March 15, 1972.

Schlessinger, F. R., et al. A Survey of Science Teaching in Public Schools of the U.S. (1971), Vol. 1: Secondary Schools. Columbus, Ohio: ERIC Information Center for Science, Mathematics and Environmental Education, 1973. (ERIC Document Reproduction Service No. ED 093 715.)

Sequential Tests of Education Progress, Series II: Teacher's Manual for Administering, Scoring, and Interpreting, Grades 7-9, 10-12. Princeton, N.J.: Educational Testing Service, 1971.

Welch, W. Personal Communication, December 7, 1978.

Welch, W. W. and A. R. Gullickson. "A Strategy for Evaluating the NSF Comprehension Program for Teacher Education." School Science and Mathematics, 73(9):759-767, 1973.

Willson, V. L. and A. M. Garibaldi. "The Association Between Teacher Participation in NSF Institutes and Student Achievement." Journal of Research in Science Teaching, 13(5):431-439, 1976.