In this seventh paper of the Research Review Series, commissioned by the Science and Mathematics Education Information Analysis Center in cooperation with the National Association for Research in Science Teaching, 189 studies in the ERIC SE collection which are concerned with secondary school science, published in 1968 or 1969, are cited and analyzed. Tables show the prime science discipline focus of the documents, the type of report analyzed, and the primary subject categories of the research. The research findings are summarized and illustrated by selected report which the reviewer considered effectively characterized the results; were clear, representative and generalizable; showed adequate methodology; or, occasionally, illustrated a particular point (good or bad). The summary is arranged under the following headings: Learning Theory and Processes; Teacher Characteristics; Teacher Education; Curriculum Development; International Education; Student Characteristics; Evaluation; Academic Achievement; Educational Programs; and Instructional Procedures. Some comments on the inadequacies of the research reviewed, and suggestions for overcoming these, are made. (AL)
RESEARCH REVIEW SERIES - SCIENCE
PAPER 7
A SUMMARY OF RESEARCH IN SCIENCE EDUCATION
FOR THE YEARS OF 1968-69
SECONDARY SCHOOL LEVEL

by
Wayne W. Welch
University of Minnesota
Minneapolis, Minnesota

THE OHIO STATE UNIVERSITY
ERIC Information Analysis Center
for Science and Mathematics Education
1460 West Lane Avenue
Columbus, Ohio 43221

June, 1971
The Science and Mathematics Education Information Reports are being developed to disseminate information concerning documents analyzed at the ERIC Information Analysis Center for Science and Mathematics Education. The reports include four types of publications. Special Bibliographies are developed to announce availability of documents in selected interest areas. These bibliographies will list most significant documents that have been published in the interest area. Guides to Resource Literature for Science and Mathematics Teachers are bibliographies that identify references for the professional growth of teachers at all levels of science and mathematics teaching. Research Reviews are issued to analyze and synthesize research related to science and mathematics education over a period of several years. The Occasional Paper Series is designed to present research reviews and discussions related to specific educational topics.

The Science and Mathematics Education Information Reports will be announced in the SMAC Newsletters as they become available.
Research Reviews are being issued to analyze and synthesize research related to the teaching and learning of science completed during a two-year period of time. These reviews are organized into three publications for each two-year cycle according to school levels--elementary school science, secondary school science, and college science.

The publications are developed in cooperation with the National Association for Research in Science Teaching. Appointed NARST committees work with staff of the ERIC Center for Science Education to evaluate, review, analyze, and report research results. It is hoped that these reviews will provide research information for development personnel, ideas for future research, and an indication of trends in research in science education.

Your comments and suggestions for this series are invited.

Stanley L. Helgeson
and
Patricia E. Blosser
Editors

Sponsored by the Educational Resources Information Center of the United States Offices of Education and The Ohio State University.

This publication was prepared pursuant to a contract with the Office of Education, United States Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their judgement in professional and technical matters. Points of view or opinions do not, therefore, necessarily represent official Office of Education position or policy.
TABLE OF CONTENTS

Part I. Nature of the Research Conducted .................. 1

Part II. Summary of Research Findings .................. 9

Teacher Characteristics .................................. 11
Teacher Education ....................................... 14
Curriculum Development ................................ 17
International Education ................................ 20
Student Characteristics ................................ 22
Evaluation ............................................... 25
Academic Achievement .................................. 30
Educational Programs ................................... 33
Instructional Procedures ................................. 34
Closing Comment ........................................ 39

Bibliography ............................................. 40

Additional References ................................... 47
REVIEW OF RESEARCH 1968-69
Secondary Level Science
by
Wayne W. Welch, University of Minnesota

A review of research is a distillation of the work of many people. This review characterizes and summarizes the research conducted in secondary school science education during the two years, 1968 and 1969. The source documents are those held by the Science and Mathematics Education Information Analysis Center (SMAC), an ERIC center. All documents were focused on the secondary level, and were published during the years 1968 or 1969. The purpose of the review is two-fold: to characterize the nature of the research carried out during this biennium, and to present a distillation of the findings.

PART I. NATURE OF THE RESEARCH CONDUCTED

A total of 204 documents were identified. The documents were of three types: (1) indexed articles from published journals, (2) dissertation abstracts (some indexed), and (3) abstracts of nonpublished reports, usually of federally funded projects. A breakdown of the documents is presented in Table 1.
TABLE 1
NUMBER AND TYPE OF DOCUMENTS
SMAC HOLDINGS IN SECONDARY LEVEL SCIENCE, 1968-69

<table>
<thead>
<tr>
<th>Type</th>
<th>Number</th>
<th>1968</th>
<th>1969</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indexed articles (mainly journal articles)</td>
<td>98</td>
<td>50</td>
<td>48</td>
</tr>
<tr>
<td>Dissertation abstracts</td>
<td>70</td>
<td>12</td>
<td>58</td>
</tr>
<tr>
<td>Abstracted report (nonpublished)</td>
<td>36</td>
<td>27</td>
<td>9</td>
</tr>
<tr>
<td>TOTAL</td>
<td>204</td>
<td>89</td>
<td>115</td>
</tr>
</tbody>
</table>

Of particular interest is the relative contribution to the total research effort made by the various types of activities. More than a third of the documents are dissertations, while one-sixth are nonjournal articles. This suggests that more than half of the total science education research effort is either an initial research undertaking (dissertation) or a nonpublished report.

Dissertation abstracts were those oriented to science education appearing during the two years 1968 and 1969. Occasionally a study was conducted prior to 1968, but was included in this review because the dissertation was published either in 1968 or 1969.

Finally, abstracted reports contained subject codes, descriptors, and a paragraph or two summarizing the content of the document. These reports tended to be either theses not carried by dissertation abstracts, or final reports of federally funded projects. On occasion there was overlap with documents listed in
the other two categories. For the purpose of characterizing the research, however, they are listed as separate documents. In such cases the article most readily available is listed in the bibliography.

Several other analyses of the research documents were conducted to help characterize the research effort of the biennium. These results are presented in the several tables that follow.

**TABLE 2**

SUBJECT MATTER LEVEL OF RESEARCH
SECONDARY LEVEL SCIENCE, 1968-69

<table>
<thead>
<tr>
<th>Level</th>
<th>Number of Studies</th>
<th>Per Cent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biology</td>
<td>47</td>
<td>23.0</td>
</tr>
<tr>
<td>Physics</td>
<td>34</td>
<td>16.7</td>
</tr>
<tr>
<td>Chemistry</td>
<td>20</td>
<td>9.8</td>
</tr>
<tr>
<td>Physical Science</td>
<td>18</td>
<td>8.8</td>
</tr>
<tr>
<td>General Science</td>
<td>16</td>
<td>7.9</td>
</tr>
<tr>
<td>Two or More Fields</td>
<td>16</td>
<td>7.9</td>
</tr>
<tr>
<td>Earth Science</td>
<td>9</td>
<td>4.4</td>
</tr>
<tr>
<td>Unified Science</td>
<td>4</td>
<td>1.9</td>
</tr>
<tr>
<td>Unclassified*</td>
<td>40</td>
<td>19.6</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>204</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>

*Means the research was not discipline oriented, e.g. study of a teacher education program.
Examination of Table 2 reveals that studies concentrating on biology appeared most often with physics a close second. Chemistry, physical science, and general science followed in that order. Twenty percent of the studies were unclassified. This term was used to describe studies that did not focus on a specific discipline, but nevertheless were science education research. Some examples include studies of teacher education programs and surveys of research.

A finding surprising to this reviewer was the small number of studies devoted to unified science. Recent activity suggests a movement in this direction. The fact that only four studies reported during a two-year period were concerned with this topic indicated little research interest in the movement.

A finding related to the results of Table 2 is the grade level at which the research was conducted. Of the 204 studies reviewed 106 were at the senior high school level, 50 were at the junior high level (grades 7-9), and 27 were concerned with teachers of science. The remaining 11 were a mixture of general surveys, studies that overlapped the junior and senior high school, and unspecified reports.

It occurred to this reviewer to examine the contribution to the research effort of the alphabet programs of the sixties (BSCS, HPP, etc.). By 1968, most programs had been in existence for some time and appeared to be making a significant impact on the research
conducted. To assess this impact quantitatively, all studies devoted to one of the newer curriculum programs were identified. The results of that analysis are presented in Table 3.

TABLE 3
SECONDARY LEVEL SCIENCE, 1968-69 STUDIES INVOLVING ALPHABET PROGRAMS

<table>
<thead>
<tr>
<th>&quot;New&quot; Curriculum</th>
<th>Number of Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSCS</td>
<td>24</td>
</tr>
<tr>
<td>HPP</td>
<td>11</td>
</tr>
<tr>
<td>CHEM</td>
<td>10</td>
</tr>
<tr>
<td>PSSC</td>
<td>7</td>
</tr>
<tr>
<td>ISCS</td>
<td>3</td>
</tr>
<tr>
<td>ESCP</td>
<td>2</td>
</tr>
<tr>
<td>IPS</td>
<td>2</td>
</tr>
<tr>
<td>Nuffield</td>
<td>2</td>
</tr>
<tr>
<td>CBA</td>
<td>1</td>
</tr>
<tr>
<td>SSSP</td>
<td>1</td>
</tr>
<tr>
<td>Combinations</td>
<td>7</td>
</tr>
<tr>
<td>TOTAL</td>
<td>70</td>
</tr>
</tbody>
</table>

As indicated in Table 3, 70 documents were devoted to one or more of the new programs. This comprises approximately one-third of the total. It is interesting to note the projects that attracted the researchers. The BSCS biology program has a history of research
and evaluation associated with its development. This coupled with the general attractiveness biology has for researchers (see Table 3) explains its relative popularity.

A surprising result is the second-place rank of Harvard Project Physics (HPP). Although the project began in 1964, the commercial version was not available until 1970. However, the project was sponsored primarily by USOE, organization more in sympathy with educational research than the National Science Foundation, prime sponsor of most of the projects. Hence, the development of the program included considerable attention to research and evaluation. A total of 11 studies appeared during the two-year period.

A final way to characterize the research of the biennium is presented in Table 4. All SMAC documents are coded according to the principal subject of the study. Although the categorization is far from perfect, it does suggest a means by which the research can be described.

The subjects of the research documents are listed according to their frequency. A description of each category will become apparent upon reading Part II of this review when summaries of the findings are presented.

The topic that was the focus of most of the research studies during the two-year period was Instructional Procedures. Nearly a fourth of all the documents were devoted to this topic. Included here were comparisons of inquiry-inductive versus expositive-deductive teaching, a topic of puzzling charm in science education.
TABLE 4  
FOCUS OF RESEARCH  
SECONDARY LEVEL SCIENCE, 1968-69

<table>
<thead>
<tr>
<th>Focus of Research</th>
<th>Number of Studies</th>
<th>Per Cent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructional Procedures</td>
<td>48</td>
<td>23.5</td>
</tr>
<tr>
<td>Evaluation</td>
<td>33</td>
<td>16.2</td>
</tr>
<tr>
<td>Curriculum Development</td>
<td>20</td>
<td>9.8</td>
</tr>
<tr>
<td>Academic Achievement</td>
<td>18</td>
<td>8.8</td>
</tr>
<tr>
<td>Teacher Characteristics</td>
<td>14</td>
<td>6.9</td>
</tr>
<tr>
<td>Student Characteristics</td>
<td>13</td>
<td>6.4</td>
</tr>
<tr>
<td>Teacher Education</td>
<td>11</td>
<td>5.4</td>
</tr>
<tr>
<td>Learning Theory and Process</td>
<td>11</td>
<td>5.4</td>
</tr>
<tr>
<td>International Education</td>
<td>9</td>
<td>4.4</td>
</tr>
<tr>
<td>Educational Programs</td>
<td>8</td>
<td>3.9</td>
</tr>
<tr>
<td>Others</td>
<td>19</td>
<td>9.3</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>204</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>

literature for several decades. Another comparison science educators often feel compelled to make is laboratory activity with lecture-demonstration activity. More recently, the popular comparison is the "new" curriculum program with the "conventional." However, according to the coding procedures adopted by SMAC, the last
comparison is more likely to appear as an Evaluation study rather than an Instructional Procedure. These comparison studies account for the relatively large number of Evaluation studies (33) listed in Table 4.

An additional point should be made about the categories listed in the table. Only ten of a possible 26 are listed because of the absence of documents in many cases. Some examples with fewer than 5 documents include Science Facilities, Educational Media, Educational Objectives, Research Methodology, and Science and Society. The table presents a brief summary of the science education research of the period. A comparison of such a table over several issues of this review would provide readers with trend data in research. Perhaps it would then be possible to take the next step, and begin to understand the factors that determine the selection of research topics.

The categories in Table 4 are the rubrics under which the distillation of the research findings of Part II will be presented. All studies falling into a given category were examined and several were selected to illustrate the general tendency of the findings within the category. Such a selection process necessarily implies judgmental decisions on the part of the reviewer. Articles for discussion were selected on the basis of their generalizability, adequacy of methodology, interest, representativeness, clarity of findings and occasionally because they illustrated a particular
point (both good and bad) that needed to be made. The purpose of Part II of this review is to summarize the findings. The articles selected for discussion were, in the opinion of the reviewer, the most effective in characterizing the results.

PART II. SUMMARY OF RESEARCH FINDINGS

Learning Theory and Processes

The unifying characteristic of the studies fitting into this category was the application of certain learning principles to the teaching of science. Pella and Triezenberg (39) found mixed results on tests of knowledge, comprehension, and application in attempting to determine the effectiveness of advanced organizers. Different levels of abstraction (verbal, pictorial, and models) used to present the advanced organizers (central unifying ideas) differentially affected the learning of seventh and ninth grade students. Working models as organizers had a significant effect on comprehension items, but not on items measuring knowledge and application. The effects of the different organizers on retention scores after a six weeks' duration are mixed. In spite of the inconclusive results, the study appears to be a model of effective research. It is based on accepted learning theory and the research design is excellent.

Another study, Mascolo (33), supported the value of organizers as an aid to learning. His results indicated that knowledge organized around the conceptual schemes has greater affective
meaning for students, and is positively related to performance in conceptualizing. He found that no significant differences could be attributed to formal training and practice in the skills of logical inquiry.

In a related study, Duncan (25) tested the influence of different types of examples in teaching concepts of genetics. He found that concrete exemplars were no more effective in attaining concepts through instruction than were verbal and symbolic exemplars. The format of the instruction was a linear program. A significant level of achievement was noted for IQ scores.

In another test of the effectiveness of various types of programmed instruction, Darnowski (20) found that immediate learning was greater for programmed instruction than for traditional lecture-demonstration techniques. The branched form resulted in significantly greater gains than did other forms of linear programs. However, the control group (lecture-demonstration group) showed significantly greater retention of facts and principles than did any of the groups using programmed materials. The results suggest a type of "cramming" effect growing out of the learning acquired from short-term programmed instruction materials.

Other factors affecting learning were investigated by Anderson (2) and by Cole and Raven (18). Anderson found that classroom social climate variables do affect individual learning, and that
climate properties affect learning differentially for various measures of learning, and for students differing in sex and mental ability. His climate variables were measured using the Learning Environment Inventory (LEI), an instrument assessing students' perception of their class.

Cole and Raven (18), in a carefully designed study, found that learning the correct principle without first learning to exclude the false principles is not as effective as learning to exclude the false principles and then learning the correct principle. Their work concerned learning the principles of floating bodies with a sample of eighth grade students.

Teacher Characteristics

Included in this category were studies of teacher verbal behavior in the classroom, teacher personality characteristics, and the effect these characteristics had on learning. Another popular topic was the development of modified classroom observation techniques and their use in determining differences among teachers using the "new" curriculum and those using the "traditional" approach.

As a general summary, one could say that teacher personality characteristics do have a significant effect on what students learn (Rothman, Welch, and Walberg (43)), and on student attitudes (Schmedemann and La Shier (48)). It was also determined in three of four studies that the mere using of a specific curriculum did
not significantly affect either the teacher's verbal behavior or his attitude (Matthews (34), Balzer (4), Petit (41)). Contrary to this general finding was a study by Vickery (63) that indicated teaching behavior can be modified by the provision of instructional materials; in this case, the Intermediate Science Curriculum Study (ISCS). In each of the above cases, teacher behavior was assessed using observational techniques.

In an interesting study by Campbell (16) involving ten junior high school teachers of low achievers, the teacher exhibiting indirect behavior, as defined by Flanders, was far more successful in terms of cognitive and affective development than was his more direct counterpart. The student criteria included measures of scientific attitudes and curiosity, and general science achievement.

Differences in teacher attitudes and personality for various selected groups were examined by several investigators. Walberg and Welch (64) found differences in personality characteristics between those teachers defined as innovative and teachers in general. Butler (15) found slight variations in graduate and undergraduate course patterns between those BSCS teachers who held positive attitudes toward the subject and those teachers who held negative attitudes toward the subject. Sadler (44) found differences in personality characteristics for those teachers with positive attitudes toward the PSSC course from those with negative attitudes toward the PSSC course. Dorsey (24) compared the characteristics of teachers that had been selected to attend NSF-AYT
institutes with those teachers who had been rejected. He found differences between the two groups, particularly on scores of the Edwards Personality Preference Schedule (EPPS). The subscores on need for Order and need for Heterosexuality were the best discriminators. This was the second time the EPPS subscore, Heterosexuality, appears in this grouping. Rothman, Walberg and Welch (43) found that teachers' scores on this test were the best predictor of physics students' achievement and attitude changes.

On the basis of the above teacher characteristics studies, it seems reasonably clear that if a select group of teachers is identified, the attitudes and personalities of this group will be different than those of some nonselect group. This finding is not too surprising, yet it is only within the last several years that this area has been studied by science education researchers. An interesting and important extension of this work is the connection some investigators have been able to make between teacher personality and student learning. It seems clear that the personality variables are the best predictors of student learning, yet our teacher training programs continue to focus on academic preparation. The attempt to select teachers for their affective impact on students is an area in need of considerable additional study.
Teacher Education

The studies devoted to teacher education were a curious mix of inservice and preservice programs, with generally disappointing findings. They were curious in the sense that innovative ideas appeared to be lacking, and for the studies lack in producing significant results.

Perhaps the study reported by Breit (12) best typifies this group. He examined the relative effectiveness of a teacher education program given at the preservice level and at the inservice level. The criteria were the development of certain teacher competencies: knowledge, positive perception of goals and methods of a curriculum innovation, and facility in coping with a learning situation that emphasizes the child's responsibility for his own learning. The subjects were 58 students enrolled in an undergraduate science methods course, and 28 elementary science teachers enrolled in an inservice version of the program.

Both versions of the program produced significant gain scores on the three criterion instruments on a pre and post-test basis. The magnitude of the gain scores varied between groups on the different measures, but this appeared to be mainly due to variations between the two groups of participants.

The problem in this study and others like it lies in the difficult task of trying to pinpoint the causal factors for the results noted. Although significant results were obtained, it is virtually impossible to ascertain what program activities "caused" the results.
Other investigators developed and/or evaluated a variety of inservice programs. In one of the few experimental studies reported, Welch and Walberg (70) found that four NSF summer institutes produced significant cognitive growth on measures of achievement and general understanding of science.

Beard (6) sought to find a relationship between teacher presentation style and the effectiveness of that teaching using the criterion of student achievement. A desired style was presented to trainees in a curriculum workshop. Ratings of videotapings of teacher presentations provided the data for the analysis. Although Beard found considerable variation in teaching style and student achievement among teachers, the workshop was not successful in influencing teaching techniques.

Using the criterion of student scores on the Portland Science Test, Stronck (52) found that an inservice assistance program was not effective. Teachers who were given only the text of a new program had students showing performances equivalent to those shown by teachers enrolled in the inservice program. Furthermore, student performance of both groups was lower than in preceding years. While one can question several methodological problems (e.g. appropriateness of the control group) the overall results are disappointing. There is the implication that we do not know what constitutes an effective inservice program.
Several studies of preservice programs were reviewed. Goldthwaite (27) used microteaching as a vehicle to develop demonstration skills prior to student teaching. Interestingly, the group that served as students for the microteaching developed the greatest proficiency in giving class demonstrations.

Uhlhorn (62) found value in providing students with teaching experience prior to student teaching. He presented evaluations of initial student teaching performance as part of their student teaching instruction. While the study has little generalizability to other situations, the techniques employed seem to hold promise.

Although there was a time when survey studies in science education were so numerous that reviewers could easily dismiss most of them, the two surveys that appeared in this collection both seemed worthy of further mention. Tweeten and Yager (61) found what they perceived as serious shortcomings in the content preparation of biology, chemistry, and physics teachers in the state of Iowa. The data were obtained from a survey of the content preparation of all Iowa teachers. The general tone of this study is "My, what poor science backgrounds these teachers possess." However, because we do not really know what background is desirable or even minimal, we cannot make such judgements. (See section on Teacher Characteristics.)

Mayer (35) surveyed the earth science teacher preparation programs of 397 institutions and found that 31 per cent had special
training programs. He also noted increases in the number of institutions offering programs and provided summary data on the nature of the various preparatory programs.

Curriculum Development

The 20 documents falling into this category are linked by their descriptions of various curriculum development activities. As such, they are not research studies in the usual sense. Occasionally the development activity is complemented by some attention to evaluation, but this is not the usual situation.

Four of the studies described the development of integrated science programs. Day (22) developed a model for secondary school physical science that included the contiguous components of the PSSC and CHEM programs. A descriptive rationale for the model was written based on progression of concepts from the simple to the complex. He assessed the validity of the model using a panel of judges.

The effects of an integrated science course of a different type were studied by Cossman (19). While this study describes an evaluation of the course in achieving the aim of fostering scientific literacy and the adoption of scientific attitudes, it is included in this grouping because its focus is on a new curriculum. The integration for this course is science with social studies. Cossman found significant gain scores on measures such as the Test on Understanding Science, Watson-Glaser Critical Thinking Appraisal, and the Study of Values. However significant growth scores were
not obtained on a standardized achievement test. In other studies, Showalter (51) evaluated a four-year sequence for unified science, while Amend (1) developed a set of laboratory blocks in nuclear physics for advanced physical science students.

Much has been written about individual learning packages or modules. However, little attention to developing such modules in science was evident during 1968 and 1969. Only two studies reviewed described such activity. Bass (5) developed and evaluated a series of three self-instructional units for use in physical science courses. Each activity was designed to lead the student from his experiences with physical apparatus to the formulation of a mathematical equation. At least one of the three series was successful in leading students to desired objectives. Certain recommended changes were suggested for the other two. In the other study, DeRose (23) described a specific and local independent study program.

Three studies related to curriculum development were concerned with a single concept or set of related concepts. Benson (7) described a programmed unit in statistics for use in science courses. Poolton (42) developed a problem-solving unit on nutrition for a junior high school science course. Finally, Shockley and others (50) discussed an approach to teaching science concepts that they called "concept distillation." Experiences, games, and puzzles containing the distilled essence of the concept of conservation of
energy were developed. A series of five booklets requiring ten hours of instructional time was written to illustrate the basic meaning of the conservation principles. Evaluation of the USOE-sponsored project suggests this approach was successful in producing changes in attitudes about science and problem solving.

Two studies focused on specific target groups of students in attempts to produce appropriate science materials. Berryman (8) designed and evaluated materials and teaching procedures that were prepared expressly for disadvantaged low achievers in eighth grade general science. While the evaluation results are less than convincing, the glimmer of activity in science for the disadvantaged is worth reporting.

In the other study, Walton (65), sought to design a special chemistry course for non-collegebound students. The text and laboratory manual were pilot tested in small rural schools in the Southwest. Although the initial low scores of the experimental group were partially due to regression effects, these students did show considerable gain over the control group on chemistry achievement tests. As mentioned earlier, it is difficult in a review of this type to ascertain what it was about the specially developed materials that would account for their apparent success. Perhaps a perusal of the complete materials would be helpful. However, limits of time do not permit such activity. It therefore seems incumbent upon the developer of new materials to describe carefully
in his article just what it is to which he attributes his apparent success. Until such responsibility is met, the science education research community remains in the dark and has little more than hope that some new key to learning has been discovered.

International Education

The main type of document included here is the doctoral thesis conducted in a foreign country by a student who is working on his degree in the United States. In fact, eight of the nine documents examined were doctoral theses. Three of the studies were devoted to surveys of science education facilities, teachers, and methods of teaching. In addition, there was one experimental study, two studies concerned with implementation of "new" programs, one on earth science course development, and finally, two relating various teacher characteristics to student learning.

Geographically, four studies were devoted to African education, two concentrated on countries in the Middle East, two were set in India and East Pakistan and finally, one took place in Australia. It appeared that the developing countries were of most interest to the science education researcher.

Studies conducted on an international scale seem to suffer the same problems of inexperience, time, and funding as their American counterparts. Most appear to be one-shot affairs conducted primarily under the pressure of finishing a doctoral thesis. Long-range and supporting studies are virtually nonexistent. In
addition to these problems, the ever present, "No significant differences were observed between the two treatments, even though..." appears as often as it does at the conclusion of American-based studies.

A study by Cannon (17) illustrates the point. He developed two different versions of a program on the topic "light"; one where an overt response was required and one where the same material was read with the response filled in. Apparently, the supplying of the obvious response was thought to have a differential effect on the learning. In testing the effectiveness of the two programs using a locally constructed criterion test, the investigator reports, "No significant difference in post-test scores on the criterion test was observed between the experimental and the control group, but the adjusted means for the direct reading group were higher than for the overt responders."

There are several problems in this kind of study, whether it be done on students in Tanzania or in Texas. The basic problem is the insinuation that because the one group scored higher than the other, something was "really" happening but we were not quite clever enough to find it with our tests. Additionally, a logical analysis of the problem suggests little hope for dramatic differences on such a minor difference in treatment.

The survey type international study is illustrated by Busch (14). He gathered data on science laboratories, teachers, service loads,
outside employment, and teaching methodology in Jordan and Syria using a short questionnaire distributed by the Ministry of Education. The findings are presented as characterizations of the typical science teacher in the two countries.

**Student Characteristics**

There was a lack of similarity among the 13 studies reported in this category. Each seemed to have a different focus and it is difficult to summarize the findings. Accordingly, a few of the better designed studies were selected for summarization.

Willoughby (72) attempted to determine the influence of certain personality factors and the type of physics course taught on the Test on Understanding Science (TOUS). He assessed the personality characteristics of nine classes of PSSC students and 13 classes of conventional physics students using the California Test of Personality. No relationship between these characteristics and student scores on the TOUS test were found. He concluded that personality is not a significant factor in student growth of understanding science.

When compared to normative data, students enrolled in physics scored lower in social adjustment. However, this finding was not related to TOUS scores. While differences on the TOUS test favored the PSSC course, these differences were not significant when TOUS pretest scores were used as a covariate.
In a similar study Welch (37) described several characteristics of students enrolled in physics. Using a nationwide sample of 2,419 students, Welch found the typical male physics student to be bright compared to his peers and possessing strong interests in mathematics and science. Such a student places high value on the pursuit of truth, is more than likely a school leader with strong power and accomplishment drives, and probably a senior planning on attending a four-year college.

The typical girl enrolled in physics is more often than not superior to her classmates in intelligence and in her general understanding of science. She also has strong interests in mathematics and science. In summary, the students enrolled in physics are among the most academically able in our schools.

In one of the few studies conducted in Canada and reported in the United States literature, Field and Cropley (26) related student cognitive style and science achievement. Using measures of mental operations, originality, flexibility, and category width (a scale that is not understandable from their article), they found significant differences among boys and girls. Boys' performance was superior to girls' on all four measures. Field and Cropley also found significant correlations between the measure of mental operations and science achievement scores. While it is not stated in the article, one could presume that the mental operations test is similar to most standardized intelligence tests. Other variables did not correlate with the general science test scores.
In a study yielding somewhat contradictory results, Troost (57) claimed no significant differences in general science achievement between students rated high on socio-economic variable and the rated low. He found pretest differences favoring the high socio-economic group ($t = 2.52$), but no corresponding differences on the post-tests. While regression effects would account for part of this finding, it seems inconceivable that a group of space scientists lecturing to these two groups of students during a summer program would result in such effects. Furthermore, for reasons that are altogether obscure, Troost also administered a spelling test as a post-test to both groups. He found the lower socio-economic group had significantly high spelling scores at the end of the program. It is questionable that a space science program would be expected to have differential effects on spelling ability.

In carefully rereading this rather puzzling article, a glimmer of understanding appeared. Both groups were carefully selected by school administrators on the basis that students must have IQ scores higher than 120, grades of B or better, and high scores on standardized achievement tests. The selection process was also such that students from center city areas, wealthy suburbs, and middle class communities were represented. Thus the finding is not a result of comparing socio-economic backgrounds, but rather of those factors that would lead to the selection of the participants in the first place.
One final point is puzzling. The appropriate analysis for a study of this type is to use analysis of variance techniques on gain scores. These can then be tested for statistical significance using standard t tests. This method is far superior to running separate t tests on pre and post-tests for the two groups as Troost did.

Finally, Bingman (10) found noticeable differences between participants in a local science fair and nonparticipants on the Differential Aptitude Test and the Kuder Preference Record-Vocational. However, no differences on these two measures and their subscales were noted between winners and nonwinners at the science fair.

Evaluation

Because of a duplication (a document appeared as both a journal article and as a thesis), there was a net total of 32 documents placed in this category. Of these, 20 were devoted to achievement outcomes. This summarizes the general focus of curriculum evaluation in science during this biennium. This is not intended to imply that this is in some way a waste of effort, but that this focus slights many other valuable outcome measures that can be used to evaluate curricula. It does, however, point up the strong achievement orientation of our research efforts in science education.

Examining a few of the noncognitive oriented studies will serve to highlight some of the different ways of evaluation.
The criterion of occupational interest was used by Saslaw (46) to evaluate a program for gifted seventh grade students in a longitudinal study. In addition to finding that these young students performed at levels comparable to tenth, eleventh, and twelfth grade biology, chemistry, and physics students, he also found the programs to be effective in achieving a significant increase in the personal-social factor of the criterion instrument. Comparisons were made against a matched group of control students.

In a study that used the criterion of course satisfaction, Welch (66) found that students expressed greater satisfaction in physics if they had made achievement gains, participated in science activities, and received higher course grades. Physics course satisfaction is negatively related to perceived course difficulty. The most interesting finding of the study is that it is not possible to relate course satisfaction with initial interest and ability in physics, but that it appears to be a function of what happens to the student in the physics course.

College grades were used by Bertram (9) and Bohenski (11) to evaluate the CHEM Study program and an advanced placement biology program, respectively. Bertram found that students who were enrolled in CHEM Study did not earn significantly different first semester college chemistry grades than did students who were enrolled in the conventional high school chemistry. He did find that the best predictors of college chemistry grades were (1) high school
science grades, (2) high school chemistry grade, (3) high school mathematics average, and (4) relative rank in graduating class.

Bohenski (1) sought to discover the relationship between the successful completion of advanced placement course in biology and later success in biology and related sciences. Although the examination of college grades did not provide support for the success of the programs, responses to a questionnaire by these students suggest a general feeling of satisfaction in the courses, e.g. 70 to 90 percent endorsed the program, stating that they would again enroll in such a course if they were back in high school.

An entirely different approach to evaluation was supplied by Dasbach (21) in a AAAS Newsletter. He provided prospective users of materials for unified science and mathematics programs information concerning availability, organization of equipment, percentage of materials that needed to be purchased locally, cost, and other miscellaneous data.

Another criterion for evaluating physics courses was used by Bridgham and Welch (13). They showed a relationship between grading practices in physics and declining enrollments. While it seems logical to expect students to shy away from "tough" courses, this study provides empirical support of the hypothesis.

Evaluation is generally defined as the gathering of data for use in educational decision making. Accordingly, it is necessary to have available instruments and techniques to gather
the information. To this end, several studies described the development of various instruments. Welch and Pella (69) developed a measure of science process knowledge, while Tannenbaum (54) developed a similar measure for use at the junior high level. Thompson and Schwarz (55) described the development of the physics achievement test of the "College Boards." Lewis (32) developed an achievement measure selecting various items from Bloom's taxonomical levels for use among English children.

In general, the findings of the various comparison evaluations were mixed. An examination of four different studies in which the newer chemistry programs were evaluated illustrates the results of current efforts.

One of the studies is the previously mentioned study by Bertram (9) in which he failed to find any differences in college chemistry grades between students who had been through the CHEMS course, and students who had taken the more conventional chemistry course. Using a test comprised of items designed to measure Evaluation behavior, the most complex of the Bloom domain, Schaff (47) found the treatment of kinetic-molecular theory in the CHEMS course to be more effective than the treatment in the traditional chemistry course. A pretest of items selected at the Knowledge level of Bloom's taxonomy was used in this analysis of covariance study. Also, the study was done on 24 classes, 12 using each of the two programs. In this
carefully designed and implemented study, Schaff concluded, "This investigation disclosed differences in the cognitive ability, Evaluation, of students given instruction in modern and conventional high school chemistry."

Also in an evaluation of CHEMS, Troxel (58) found students taking the CHEMS and CBA courses performed higher on a test of general chemistry achievement, the Test on Understanding Science, and the Watson-Glaser Test of Critical Thinking Ability than did students in conventional chemistry courses. Students in the conventional course and CHEMS tended to rate chemistry higher in terms of preference than did students in CBA. This study also involved the use of analysis of covariance, with the antecedent variables being each of the above instruments given as a pretest.

Finally, Osborn (37) noted the impact of the newer chemistry curricula on the criteria of changes noted on the content of the New York Regents examination in chemistry during the past several years.

The above studies were selected because of expected contradictory results, and because they approached the problem of evaluation from different points of view. However, one is led to reach a different conclusion. Granted, the CHEMS course did not result in long-range effects as measured by college grades. But the studies by Troxel and Schaff lend support to the effectiveness of the newer curricula on the selected criteria. While the sample
of four is small, it is encouraging to find a glimmer of consistent results in the evaluation of the newer science education programs.

**Academic Achievement**

Academic achievement is a rather broad category in which to group science education research studies. It is similar to educational evaluation because of the emphasis on achievement outcomes (see Evaluation section). Studies reported here are also similar to those found in the section entitled Learning Theory and Processes. However, the distinguishing characteristic seems to be a matter of focus. Whereas evaluation emphasizes the information that can be obtained for making decisions about a program, the studies in this section focus on determining what factors seem to be related to achievement growth.

Perhaps the study by Gray, et al. (28) typifies this group. They sought to determine the effect of instruction in ninth grade physical science upon achievement in high school biology. Notice that the focus is upon biology achievement, not an evaluation of the physical science course. The results are interesting. On the criterion of the Nelson Biology Test for non-BSCS students, those who had completed a previous physical science course did not achieve significantly better in biology than did students who had not completed a previous physical science course. Gray did find significant differences across the various ability groups in biology growth; that is, the high ability group gained more than the middle and the low.
In a follow-up of the data, Gray then subjected the findings to an analysis of covariance with IQ as a control. As one would expect, there still were no differences in the two groups. Additionally, no differences were found among the high, middle, and low ability students. While carefully limiting their conclusions because the study was done in one community with only four high schools, they raise serious questions about the prerequisite of physical science for biology on the basis of achievement.

Another study representative of this category was reported by Tweeten (60). He related various program characteristics to science achievement as measured by the science test of the Iowa Tests of Educational Development. The data were based on 431 schools in the state of Iowa. Tweeten found three of ten school variables related to science achievement; percentage of students in the school enrolled in science, number and variety of the new curricula being used, whether or not the school was accredited by the various accrediting associations.

Variables not related to achievement are also of interest. These included number of class preparations of science teachers, teacher academic preparation, attendance of teachers at NSF institutes.

The authors and the reviewer must be quick to recognize the multitude of variables operating within a study of this magnitude, but the findings do offer data regarding a number of cherished
beliefs in the teacher training profession. Researchers would be advised to consider some of these results when planning other studies.

A brief sampling of several other studies follows. Welch and Bridgham (68) found no correlation between physics achievement and the length of time students had studied a given unit. The criteria were regression adjusted gain scores, and the treatment duration varied from 24 to 62 days.

Swartney (53) identified several science concepts and mathematical skills required for learning chemistry. Hug (29) compared the relative effectiveness of the four components of flexible modular scheduling on achievement in biology. Trinklein (56) found short excerpts of a film as a review of the total film to be more effective than the excerpts or the film by themselves. Perkes (40) found the pattern of behavior in the class to be related to student achievement. He did not find a relationship between the number of credit hours of preparation of the teacher and student achievement in junior high school.

Atwood (3) sought to find the impact of instruction on changes in cognitive preference among high school physics students. While no significant gains were noted on a test of cognitive preference based on chemistry as a result of instruction in physics, the study is interesting in its attempt to create a content-free cognitive style test to determine the impact of instruction.
Educational Programs

Among the eight studies of this group only one met the established criteria for inclusion stated in Part I of this review. While a study of metric system instruction such as that conducted by Murphy and Polzin (36) may have interest because of the emerging developments in the adoption of the metric system, their results are far from generalizable to any other school system.

"Educational Programs" studies were a mixture, from the study of teaching about the metric system (Murphy and Polzin) to the attempt to compare curriculum project objectives against state adopted objectives (James (30)). If there is a common thread among the documents, it appears to be the concentration on either the development or analysis of a locally-developed program.

The study by Sadowski (45) is illustrative. He noted what to him was a disturbing decrease in interest in the science fairs in Genesee County, Michigan, during the period 1957-1967. This loss of interest was based on numbers of participants, particularly at the secondary level, and his observations that the quality of exhibits is also on the decline. This study typifies a type of local research all too common in the science education research journals.

Perhaps the only study of this section that does seem valid to report in our research journals is one conducted by Troxel and Yager (59). Drawing from their massive source of data on science
in Iowa schools, they have noted a steady decline in enrollment in science courses, when adjustments are made for increased school enrollments. In 1958-59, 59 per cent of the 9-12 grade students were enrolled in some kind of science course. The percentage for 1966-67 was 52 per cent. In spite of new programs and much federal money provided for schools, the appeal of science among our young people does not seem to have increased.

Instructional Procedures

If one were to summarize the results of the studies reported here, it can best be done by quoting (anonymously) the following statement, "Although the experimental group realized a numerically greater increase in scores on the biology achievement test than did the control group, the difference between the two groups was not statistically significant." This summary can be supported by examining in some detail the 48 studies.

Among the 48 documents, there was some duplication. In one case, this meant that a single study appeared four different times as a dissertation abstract, a journal article, a speech at a convention, and as an SMAC annotation. However, it was all the same summary of research. Another article appeared in three different forms, and two appeared twice each. This reduced the total numbers of separate documents to 43. In addition, five studies have previously been discussed elsewhere. Thus we are left with a net total of 38 documents.
Of these 38 studies, 30 were devoted to a comparison of various instructional methods; for example, lecture-expository versus guided discovery or laboratory versus demonstration. Table 5 reports a summary of the findings of these 30 studies.

<table>
<thead>
<tr>
<th>Finding</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>No significant differences</td>
<td>17</td>
</tr>
<tr>
<td>Mixed results</td>
<td>6</td>
</tr>
<tr>
<td>Favored experimental</td>
<td>6</td>
</tr>
<tr>
<td>Favored control</td>
<td>1</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>30</strong></td>
</tr>
</tbody>
</table>

The results of this table tell us that variation in instructional procedures of the type that science educators are usually concerned about has very little effect on the variables we are presently able to measure. Only 20 per cent of the total clearly suggested an advantage to the experimental treatment, and even in these the tone of the author's report suggests strong investigator bias.

Initially, one is moved to offer an explanation involving the failure of our measuring instruments. However, here is one area where some improvement has been shown since the days of the
Curtis reviews. No longer does the doctoral candidate rely only on an achievement test. In nearly all of the above studies, a multitude of criterion measures were used. Some of the more common were the Test on Understanding Science, the Watson-Glaser Critical Thinking Appraisal, locally-developed achievement tests, measures of attitude, lab skill tests, and questionnaires.

If any explanation can be attributed to the nature of the tests, it appears to be in the lack of connection between the instructional procedure and the test chosen to measure the effect. The TOUS enjoys widespread popularity, but as an example one doubts if TOUS measures the kind of outcome expected from a five-hour treatment of the fundamentals of unit and dimensional analysis (Williams (71)). Examinations of TOUS items clearly indicate that it does not measure all the outcomes investigators hope to achieve with their instructional panaceas.

While it can be argued that finding no significant differences is in itself important, one begins to wonder how often this argument can be used. In this review 80 per cent of the studies could not be supported by such an argument. For the benefit of the reader, a sample of focus of some of these nonsignificant findings is listed below:

Open ended versus directed laboratory
Independent study versus regular class instruction
Self-directed versus teacher directed
Team teaching versus conventional
Expository-deductive versus discovery-inductive

Variation in the format of laboratory reports

Programmed problem-solving versus nonprogrammed problem-solving

Detail in visual presentations

Teacher described demonstration versus discussion type demonstration.

After reading summaries of these articles, it appears that the problem might be the strong convictions the investigator has prior to beginning his study. Because he was so very certain that what he was about to do would drastically affect student learning, he does not bother to carefully choose his criteria to represent accurately what he expects to happen. Perhaps we should try out new procedures with a mind set that predicts failure. Then we will more carefully select the criteria that might detect the impact. Those who have done many pre-treatment-post studies know that effects are not impressive. Even the most significant (statistically) gain will probably be less than one-half of a standard deviation. This may mean two or three points of a 40-item test. This shattering realization should lead to more carefully designing the study in the first place.

Among those few studies that seemed to suggest an effect because of a selected instructional procedure, several are worth mentioning. The studies by Pella and Sherman (38) and by Yager, et al. (73) illustrate the points of the preceding paragraph. In both carefully designed studies examining several outcome variables,
the results are consistent with expectations. Pella and Sherman found that pupil contact with apparatus was more effective than noncontact on a measure of laboratory skill. However, no significant differences were noted on several cognitive and interest measures.

Similarly, in comparing laboratory and demonstration techniques, in what should probably be the last study of this type, Yager, et al. (73) found the laboratory to be more effective on a test of practical laboratory skills. Again, no significant results were noted on several other cognitive measures.

Schuck (49) found that training in set induction techniques was very successful in developing pupil achievement and pupil perception of effective teaching. The purpose of set induction is to focus pupil attention on some common referent which becomes the vehicle by which the teacher makes the transition to new material and builds continuity from lesson to lesson.

Lauren (31) found that varying teacher roles in the classroom produced significant effects on students' perception of classroom climate, and on the teachers' indirect versus direct behaviors. Indirect teaching results in more positive student perception and increased achievement in earth science. The study has greater importance because it sought to find relationships among a group of low-achieving students.
The remainder of the studies reporting significant findings are not reported because of their local focus, poor design, or contradictory findings.

The eight studies that were not included in Table 5, but were found in the Instructional Procedure category were of two kinds: description of a procedure with no mention of its effectiveness or general surveys of program usage, e.g. kinds of activities taking place in schools using the "new" curricula.

Closing Comment

As mentioned initially, this is a review of research in secondary school science education for the biennium, 1968-1969. It is a distillation and characterization of the research efforts of science educators. Specific references have been pointed out in the various sections and several documents have been used to illustrate the points. Much of the work appears fragmented and uninspiring with little impact in the schools. Perhaps this is due to the fact that motivation for most of the work is completing some external requirement, a thesis or funding report, rather than generating knowledge about the educational process. This hypothesis is supported in some sense by the quality of work that appears in a few places that have ongoing research projects associated with them. There is hope for educational research only when experienced researchers are able to devote the majority of their time to seeking the answers to important educational questions. At present, we are far from that goal.
BIBLIOGRAPHY

The citations containing ED information are available from EDRS. P.O. Drawer Q, Bethesda, Maryland, 20014.


ADDITIONAL REFERENCES


ED 028 088 MF $0.65 HC $3.29 19pp


ED 020 134 MF $0.65 HC $3.29 33pp


ED 027 209 MF $0.65 HC $3.29 14pp


ED 033 854 MF $0.65 HC $3.29 9pp


ED 028 087 MF $0.65 HC $3.29 14pp


149. Newman, Harold. "Initiating a Science Reading Program in a Ghetto High School: Role of the Science Reading

ED 033 863 MF $0.65 HC $3.29 25pp


168. Shinpoch, John Russell. "A Status Study of Biology Teachers in the Public Secondary Schools of the State


178. Torop, William. "An Experimental Study to Determine the Relative Effectiveness of Four Methods of Laboratory Reports in the Teaching of CHEM Study Chemistry."


