The editors of this review have focused on four purposes: (1) to portray the state of knowledge in science education, (2) to describe any existing trends, (3) to identify areas which need to be researched, and (4) to provide tentative answers to pertinent problems, if any seem to emerge from the research. Research studies reviewed have been divided into the main categories of learning; education, characteristics and behaviors of teachers; values and philosophy; and surveys. Also, an index and bibliography are provided in order that the reader may make easy reference to the 307 research studies listed. (CP)

Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC makes every effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this affects the quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (EDRS). EDRS is not responsible for the quality of the original document. Reproductions supplied by EDRS are the best that can be made from the original.
A SUMMARY OF RESEARCH IN SCIENCE EDUCATION-1973

EDITED BY
MARY BUDD ROWE
and
LINDA DeTURE

University of Florida
Gainesville, Florida 32511
PUBLISHER’S NOTE

With this volume, responsibility for the annual publication of A Summary of Research in Science Education is being accepted by John Wiley & Sons at the request of the dedicated educators, researchers and institutions heretofore concerned with its issuance. Such action of transfer was approved by both the National Association for Research in Science Teaching (NARST) and The Association for the Education of Teachers in Science (AETS). Previous volumes, through 1972, have been produced and distributed by The Center for Science and Mathematics Education, Ohio State University. This 1973 edition and successive editions will be published in conjunction with the distinguished journal, SCIENCE EDUCATION.

SCIENCE EDUCATION, now in its 59th volume year, is expanding its scope and audience under the direction of Editor N.E. Bingham. As a result of the sponsorship of The Association for the Education of Teachers in Science (AETS), a new section, “Science Teacher Education,” has been added to the journal.

The publisher welcomes inquiries regarding subscriptions to this refereed international journal of researched practices; issues and trends in science instruction, learning, and preparation of science teachers.
Foreword

Research Reviews are being issued to analyze and synthesize research related to the teaching and learning of science completed during a one-year period of time. These reviews are developed in cooperation with the National Association for Research in Science Teaching. Appointed NARST committees work with staff of the ERIC Science, Mathematics, and Environmental Education Information Analysis Center 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
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Purposes Served by the Review</td>
<td>1</td>
</tr>
<tr>
<td>Organization</td>
<td>1</td>
</tr>
<tr>
<td>Choice of Studies and Methods for Their Review</td>
<td>2</td>
</tr>
<tr>
<td>Other Reviews and Summaries</td>
<td>3</td>
</tr>
<tr>
<td>Learning</td>
<td>5</td>
</tr>
<tr>
<td>Introduction</td>
<td>5</td>
</tr>
<tr>
<td>Learning Studies Based on Theories Other than Piaget's</td>
<td>5</td>
</tr>
<tr>
<td>Piagetian-Based Studies</td>
<td>12</td>
</tr>
<tr>
<td>Behavioral Objectives</td>
<td>14</td>
</tr>
<tr>
<td>Formats for Instruction</td>
<td>24</td>
</tr>
<tr>
<td>Curriculum Evaluation</td>
<td>27</td>
</tr>
<tr>
<td>Tests</td>
<td>32</td>
</tr>
<tr>
<td>Physics</td>
<td>36</td>
</tr>
<tr>
<td>Science and the Handicapped</td>
<td>39</td>
</tr>
<tr>
<td>Education, Characteristics and Behaviors of Teachers</td>
<td>43</td>
</tr>
<tr>
<td>Teacher Education</td>
<td>43</td>
</tr>
<tr>
<td>Teacher Verbal Characteristics</td>
<td>47</td>
</tr>
<tr>
<td>Teacher Attitudes and Values</td>
<td>50</td>
</tr>
<tr>
<td>Values and Philosophy</td>
<td>52</td>
</tr>
<tr>
<td>Surveys</td>
<td>54</td>
</tr>
<tr>
<td>Summary and Conclusions</td>
<td>57</td>
</tr>
<tr>
<td>Index</td>
<td>61</td>
</tr>
<tr>
<td>References</td>
<td>63</td>
</tr>
</tbody>
</table>
A Summary of Research
In Science Education — 1973

MARY BUDD ROWE and LINDA DeTURE
University of Florida
Gainesville, Florida 32511

Introduction

Purposes Served by the Review

In preparing this review we tried to think of what users readers
might wish to make of it. We identified four purposes the review might
serve:

1. It should portray the state of knowledge in science
   education.

2. It should describe any existing trends.

3. It should identify areas which need to be researched.

4. It should provide tentative answers to persistent
   problems, if any seem to emerge from the research.

With the bibliography and index, which is provided this year for
the first time, a reader can locate all the studies that relate to an
area of interest to him. If the organization we chose does not meet
the needs of a particular reader, he can resort to making his own with
the aid of the indexes.

Organization

The organization we chose is empirically derived. It takes its
form from the studies with which we had to deal. We begin with learn-
ing research as the base. From this we proceed to educational objec-
tives and what they are supposed to accomplish. The tie between advanced
organizers and objectives seemed to us obvious but the connection was
not so apparent to the people who did these investigations. Next we
consider various modes for presenting information to people. Here we
operate on the thesis that people exhibit different patterns of apti-
tudes. These patterns imply that a given treatment or mode of instruc-
tion will not be equally suitable for all people. People will be
differentially sensitive to information presented in audio or visual
modes, for example, or in self-paced as opposed to fixed-paced instruc-
tion. More investigators are doing the kind of aptitude-treatment
interaction studies that will allow us to identify which treatments are
appropriate for which students.
Only one subject has been singled out for special consideration, physics. The persistent decline in enrollment prompted us to put all of the physics-related research in one bundle to see whether it yielded any helpful information.

In the realm of teacher education there was a great deal of attention given to the impact of training on attitudes but considerably less effort went into finding out whether the changed attitudes resulted in changed teaching patterns. Interaction studies continue to be produced with variables such as pacing and questioning as a focus. Some work has gone into study of teacher and student perceptions of each other, as well as the purposes to be met through their joint immersion in science. Every year there are a few studies that describe the personality characteristics of some group of science teachers but the meaning or usefulness of this information receives little consideration. No one has let us in on why any particular combination of characteristics is especially useful or worthwhile in science education.

Probably the research on handicapped children and their response to science has the most clear implications for practice. Early and extensive exposure to an activity-based science program appears to be especially useful for both physically and socially handicapped children. These are the groups which seem to gain most benefit from early exposure to an activity-oriented science program but these are also the groups which rarely receive science instruction.

The largest body of research revolves around audio-tutorial or self-paced instruction. Some investigators spoke of individualized instruction but their programs were only self-paced. If a program is individualized, it is not only self-paced, it makes the materials available in more than one presentation format. From the research in this area we learn that students procrastinate; that a diet of all one kind of presentation produces educational indigestion; that students learn more and suffer less from procrastination if the information occurs in short or brief rather than lengthy units. Frequent short tests rather than few long tests spaced at longer intervals produce better achievement, especially for middle and low ability students.

We found only three studies that dealt with philosophy and values in scientific inquiry.

Only those surveys which have at least a quasi-experimental character to them or surveyed a large segment of the population are reviewed. The rest may serve specific local interests but appeared to contribute little to general knowledge or understanding about education in science.

**Choice of Studies and Methods for Their Review**

We had access to about 350 abstracts of dissertations and journal articles. In more than half the cases we found the abstracts inadequate for our purposes. It may be useful to people who plan to publish research to see the kind of questions we wanted answered in an abstract.
1. What are the independent variables?
2. What are the dependent or outcome variables?
3. How are the variables measured?
4. Who are the subjects of the study?
5. What statistical procedures are employed?
6. What are the main findings?

We regarded answers to these as the minimum information necessary to make a judgment about a paper. Many abstracts failed to supply this minimum information. In the case of journal articles we read the article. This was not practical to do for the dissertations. The latter were assessed solely on the basis of the abstract.

We have one other suggestion or comment to make concerning tests of significance and costs of a treatment. A statistically significant outcome may not be a functionally significant outcome. Most investigators failed to say how much difference between treatment groups would be enough to make a treatment worthwhile. More attention should be given to the power of a test of differences. An alarmingly large number of people referred to the results of a study as positive, i.e., in the expected or desired direction, but non-significant. If a result is not significant, one cannot say anything meaningful about the direction of the difference. The inference is that, if the sample were larger, things would surely turn out in the direction the investigator expected. That notion is erroneous and ought to be discouraged both because it is bad statistically and because if the difference in means is small but significant we still must ask, "Is the difference functionally significant?"

Other Reviews and Summaries

Wall (295) reviewed 58 studies in astronomy education. Knight (141) surveyed all the dissertations in science and mathematics in Vol. 31 of Dissertation Abstracts.

Tisher (277) edited a monograph that contains 10 articles presented at the Third Annual Conference of the Australian Science Education Research Association.

Balzer et al (23) did a review of research on teacher behavior as it relates to science education. It was published as the first Yearbook of the Association for the Education of Teachers of Science (AETS). Section one discusses instruments and their development. Section two describes research on teaching published in the period from 1960-1971. The third section gives some perspective to the science research through discussion of teacher behavior studies in other fields. They conclude that teachers dominate classroom talk and students seldom initiate it.

We also want to call attention to Report 7, on Science, in the National Assessment of Educational Progress (196). This report is
based on the 1969-70 survey designed to measure scientific knowledge of children and adults. Blacks performed between 11.8 and 15.8 percent below the national average at all four age levels tested (9, 13, 17, 26-35). With adjustments for various background variables the residual disadvantages fell to 7.7 percent. Blacks performed best on those science exercises which depend mostly upon daily experience and common knowledge and poorest on those which involved a detached research attitude toward objects and phenomena. Interpretations of the National Assessment findings in science appeared in Science and Children, Vol. 11, No. 1, 1973; the Science Teacher, Vol. 40, No. 6, 1973; and the Journal of College Science Teaching, Vol. 3, No. 1, 1973. Males performed better than did females. Inner-city students showed a large deficit.

Comber and Keeves (55) gave an account of the conceptualization, execution, analysis and results of the international science study which involved 19 countries. Bridgham (36) reported on the National Longitudinal Study in Mathematics. He discussed the results of attempts to correlate mathematics with achievement in science.

Strassenberg and Paldy (264) summarized the main funding thrusts of nine foundations so that researchers in science education might have a better idea where to place proposals. The Carnegie Foundation supports research and development in early childhood, elementary and secondary and medical education. It seeks to promote advancement and diffusion of knowledge. The Danforth Foundation is interested in furthering development of people and values. Exxon is generous to scientists and, while it gives support to chemical engineering, it has funded some small curriculum development activities. The Kettering Foundation has supported individually guided education at elementary and secondary levels. The Rockefeller Foundation has an international thrust. Agriculture, food sciences, and other ventures meant to alleviate hunger are of interest to this foundation. The Shell Companies Foundation has supported small ventures in many areas of science education. For basic research in science, the Alfred P. Sloan Foundation may be the best prospective source of support.

Mayer (177) discussed legal ramifications of the inclusion of evolution in the curriculum. Some court cases and legislative bills have been filed to legalize the use of the creationist view. The law has not yet been accepted to the religious conception.

At the 1973 program of the National Association for Research in Science Teaching, Ralph Tyler (284) presented his summary analysis of 50 investigations completed in the period 1966-73 and compared these with research completed during 1955-65. Essentially his paper called for increasing the practical application of science research and curriculum in the lives of students. The large number of unmotivated students was the chief object of Tyler's concern. He maintained that we need to fit the objects and ideas of science into the students' world. He maintained that we need more understanding of how to join up out-of-school experiences with in-school experience. Students' purposes, habits, drives and means of achieving psychological equilibrium require more study.
We would comment that the kinds of questions one asks depends partly upon the basic model of learning he accepts. In contrast to the drive reduction, equilibrium model that Tyler uses, Bertalanffy speaks of deliberate initiation by organisms of disequilibrium - curiosity, the search for novelty, escape for boredom. This is what Bertalanffy maintains we ought to research. How can we use and nurture these propensities?

Learning

Introduction

In the NARST - ERIC/SMEAC 1972 Summary of Research in Science Education, Joseph Novak complained that too many researchers who examined achievement or learning variables failed to connect their work with learning theory. He noted that since 1920 most investigations which focused on the impact of different instructional regimes resulted in no significant differences (p. 8). Such is also generally the case for those reported in 1973. But we cannot entirely agree with Novak's contention that more steeping in learning theory might result in more productive findings. While we fully support his feeling that some fundamental paradigms ought to underlie research (p. 4), we find ourselves more in accord with Bertalanffy, the systems biologist, whose indictment of psychological research has merit. He characterizes learning research as "ratified", that is, concentrated until recently on rat derived models which fail to reflect what is unique about man-kind, namely his ability to manipulate symbols, to create meaning, to influence the quality and character of his life, and to tie the past to the future. What proportion of psychological research, he asks, has been addressed to meaningful learning? He remarks that a large part of modern psychology... covers the triviality of its results and ideas with a preposurous language bearing no resemblance either to normal English or normal Scientific Theory... " (p. 6). As we read the papers in this review we asked, "Are we equally guilty?" We leave our readers to decide for themselves.

Learning Studies Based on Theories Other Than Piaget's

With an application of Ausubel's theory of meaningful verbal learning in mind, C. I. Hall (97) tried to determine whether the use of extra exemplars in elementary reading material would improve reading comprehension. She chose three topics and provided three levels of amplification for each. How useful such exemplars are depends on what relevant concepts children already have. With a Cloze test as criterion, Hall found that no consistent pattern emerged as a result of the treatment. In some instances moderate amplification helped and in others it hindered. In individual interviews, Hall found that the 135 children

2Ibid.
in the study varied tremendously in the relevant conceptual props they already had.

If children are subject to some kind of experimental treatment, then it behooves the investigator who wants to understand why he got particular outcomes to know something about the kinds of aptitudes which characterize the people in this sample. Ukens (285), who did a study of achievement in an elementary science program, the Conceptually Oriented Program in Elementary Science (COPES), based his design on Guilford's structure of intellect theory. He attempted to find out which abilities or aptitudes were particularly related to pupil achievement in the COPES Mechanical Energy Sequence. With scores on 17 selected abilities for the 458 sixth graders in his study and results of pre- and post-tests as data, Ukens performed step-wise multiple regression on the data with the post-test scores as the dependent variable. Convergent production of figural transformations (CFT) and divergent production of semantic transformation (DST) proved to be significant predictors of post-test scores.

E. Keller (135) used the same design to examine what abilities were selectively predictive of success in the COPES Conservation of Energy Sequence. The set of abilities he found to be good predictors were convergent production of semantic relations (CSR), convergent production of figural transformations (CFT), divergent production of semantic transformations (DST), cognitive semantic relations (CogSR) and cognitive semantic classes. Thus for the two COPES units, two abilities were drawn on in common: CFT and DST.

These two studies followed one by Sher in 1971 who got similar results for a segment of COPES content. It is a line of research, sponsored by Professor Phillip Merrifield at New York University, which ought to be taken up by other investigators. Presumably different curricula, and even different units within a curriculum, draw on different combinations of aptitudes. Alternate modes of presentation may be necessary for students whose aptitude configurations are not congruent with those that predict success for that unit or curriculum. The Korans (145) report similar findings. They examined the interaction between a set of aptitude variables and three differently structured advance organizers in a programmed learning format. Their data suggest that low ability students rely more on the provided organization than do other ability groups.

Gene Moser (191) and his students at the University of Pittsburgh have been examining how information theory may be used to study human learning. The sequence of moves made by subjects in the course of learning or problem solving are considered to have Markov properties. In a test of Moser's seven-component Information Theoretic Model, Fazio (76) presented electric circuit tasks to 243 college non-science majors. Predicted random moves to solution were compared with the actual data generated by the students. The residuals were used to test

twelve hypotheses concerning the model. The procedure allows the investigator to determine the degree of structure in the search and this, Fazio found, predicts success on the tasks. Training produced a change toward more structure and, consequently, more learning. Fazio also examined conditions which increased the probability of transfer and found that a combination of exposure to more exemplars plus matching increased transfer of learning from one condition to another. Presumably the additional training influences the structure variable, i.e., teaches a productive search pattern.

D. Kuhn (149) also was interested in the impact of information density on problem solving of female college freshmen and sophomores. Based on Novak's cybernetic model, he identified three levels of analytic ability by means of a problem solving test. Then he attempted to relate information density and analytic ability to information processing output. In contrast to the Hill study described earlier, all three groups benefited from the information but high analytically benefited more. Based on Novak's model, Kuhn surmised that low analytic groups may have learned by rote and acquired fewer structures within which to incorporate information.

In passing, we might comment on points that need to be clarified concerning information density. Technically it is not just more information per page of instruction; something is information-rich according to how many new ideas are introduced. The senior author once analyzed biology texts intended for slow learners and discovered that the routine way to produce them was to remove most of the examples meant to illustrate each concept. As a result the number of new concepts introduced per page was enormous. The procedure had produced an information-dense text whose difficulty level was greatly elevated.

The Moser and Fazio work suggests how organization imposed on a task by a learner influences what is learned and how well it is retained. We can also ask, as Gagné has suggested, how structuring knowledge on a topic into a learning hierarchy shapes learning. Parker (206) suggested that adults' hierarchies and children's hierarchies might be very different. One-hundred sixth graders who were familiar with certain tasks did a task analysis to produce what seemed to them the natural sequence of ideas. Parker compared those with an adult-generated hierarchy over the same material; the Gagne hierarchy. Another group of students then received instruction from either the children's hierarchy, the adult plan, or a randomized sequence. Neither on the immediate post-test nor on the delayed post-test was there a significant differential effect on learning. Here, as in other studies, especially where the material is not especially complex, we can infer that the learner imposes his own organization on the content. It would have been a nice additional feature of this study if Parker had been able to identify attributes of those students who did well and those who did poorly in each treatment condition.

If special aptitudes or knowledge conditions performance in a given subject area, it behooves us to know in what way. In genetics, for example, probability and combinatorial concepts are central to an understanding of various genetic models. Emery (74) examined the
statistical strategies employed by a group of high school biology students confronted with certain tasks that required prediction of relationships or outcomes and identified three distinctive strategies. She did not relate these strategies to performance in the genetics course. That is, she did not regard them as possibly differential facilitating abilities. She did show that instruction in genetics did not change either the probability or the combinatorial reasoning. We could consider her to do a post hoc analysis that would involve a stepwise multiple regression analysis with scores on the genetics test as the dependent variable to determine how much variance in outcome could be accounted for by the probability strategies and combinatorial skills.

Bahorik (21) also took an interest in improving the efficiency of learning in genetics. He based his research on Snyder's prototeory of instructional efficiency. He examined the mean number of transactions needed by a teacher-student dyad to move through various stages of a problem. Bahorik also examined the effect of two variables presumably related to efficient learning: number of information items and amount of learning structure within the problem. He assumed that the least structure and the lowest number of information items would produce the most efficient task group. Sixth-graders (N = 120) were assigned to four task groups which were treated differently on the two variables described above. In practice some groups got hints, others did not. Some groups were required to read selected information while others groups had choices of what to read. The treatments failed to produce significant performance differences. However, Bahorik did confirm Snyder's hypothesis that where the ratio of teacher-student transactions approaches 1, instructional efficiency approaches a maximum.

Clarke (50) reported an attempt to apply Ausubel's Cognitive Field Learning Theory to evaluation of an Australian Science Education Project (ASEP). He rewrote a portion of the ASEP to include expository and comparative organizers. The materials also were arranged to reflect the theory in the progressiye differentiation of the content and in the integrative reconciliation of the parts. Means for the reconstructed ASEP unit were significantly better than those for the standard format. In this study we see a conscious attempt to apply theory to the improvement of a set of instructional materials followed by a test of the theory. This is a kind of research which ought to be encouraged.

In a two-part study meant to test Piaget's theory of time sense development, Mori and Tadang (188, 189) exposed students of different ages to high and low speed motion pictures. In part I they reported how the film rate was related to the estimated time it took for a body to fall. They showed activities such as rope skipping, leap frogging and ball bouncing. They found that estimates of time were much more varied for kindergarten than for elementary children. Eleven year olds deviated more on slow films than on fast but generally showed little change from one speed to another. The investigators concluded that accuracy of judgment depends on age and experiences in visual perception.

In the second portion of the study they flashed two moving pictures on the screen simultaneously and examined how different combinations of speed influenced judgments of distance and synchrony. Kindergarten
children thought slow runners went farther but elementary children recognized that they covered the same distance.

In science experiments where children must make comparisons between objects moving at different speeds (e.g., pendulums or cylinders rolling down a ramp), there may be a confounding of distance, duration and synchrony. For example, if two objects roll down a slope and one takes longer, young children may also think it went farther. Concepts whose evolution hinges on development of a stable time sense probably ought not to be introduced very early or if they are, some check should be made of the status of student perception of time.

In an interesting study of the effects of generating hunches upon subsequent search activities in problem situations, Wilson (298) assigned each of 45 students, 9-11 year olds, to one of three groups. Each group observed what was surmised to be a discrepant event or contradictory stimulus. Group 1 wrote hunches concerning the explanation. Group 2 read a set of hunches. Group 3 did neither of these things. Then all three groups were asked to classify a set of procedures for resolving the conflict and scores were obtained. Group 1, which had to generate hunches, classified fewer procedures as useful, spent more time in search, and gave higher quality solutions. There were no differences between the group that read hunches and the control group which did nothing about hunches. We may look at this study in the context of mediation theory and ask whether the requirement to form hunches acts like an organizer. Can we assume that simply reading a set of possible hunches either acts as a distractor or that their content is ignored just as instructional objectives often seem to be?

Golshan (88) based her three-month clinical study of one 11 year old girl's thought processes concerning physical phenomena on deep structure field theory (d.s.f.). The girl's thought processes could not be interpreted in terms of standard science paradigms concerning physical phenomena. She nevertheless exhibited a coherent, self-consistent and interwoven framework that governed her explanations. To examine the girl's d.s.f. Golshan developed a coding system that sorted the properties of objects and actions into what she called cognitive relational categories (c.r.c.'s). Each c.r.c. that was found to apply to the protocols was examined to see how it fitted into deep structure. She found systematic regularities among the c.r.c.'s that lent support to the notion of an undergirding deep structure. This study follows the work of J. Easley who has maintained that a return to clinical analysis for the express purpose of determining how individual deep structure functions in various disciplines will ultimately lead to presentation of content and forms of instruction more suited to the way children think at different stages of development.

In a fascinating study Jerry Horn (115) investigated the relationships between achievement motivation, information status and group discussion arrangements to risk-taking behaviors in high school biology. He used the Extremity-Confidence of Hypothesis Test and the Risk-Taking Verbal Observation Scale. Risk-Taking is apparently a multidimensional trait. Usually risk-taking by groups is greater than by the individuals in them. We call attention to this study because it focuses on underlying factors that may influence the ways in which students respond to
inquiry-based courses. These courses place students under the necessity to make predictions and to live with uncertainty respecting the ultimate correctness of their efforts at inquiry. Students differ in their risk-taking propensities and in their fear of being wrong.

Risk-taking seems to be affected mostly by the amount of information a person or group has concerning the state of knowledge about the problem. If a person believes that the answer is really not known, then risk-taking is higher than if he knows an accepted explanation exists. Horn reported that the composition of the groups must be considered. He formed like-high, like-low and unlike groups based on the motivation scores. Like groups, both high and low motivation, took greater risks than did mixed groups. Groups usually took more risks than individuals.

Lefkowitz (158) sought to determine the relationship between a personality trait (dependence proneness) and performance in a high school chemistry laboratory. He also tried grouping students on this variable and examining performance of groups in the laboratory. There was a small but significant negative correlation between dependence proneness and laboratory performance. Group membership was not a significant variable. Scores in ninth-year algebra correlated more strongly with success than did mathematical aptitude or the personality variable. This is another study which would have produced more useful information had the investigator performed a regression analysis. Then we would know how much variance on the outcome measure was associated with each variable in the study.

In a five-year study, Scott (237) found that students in an inquiry type of program exhibited more analytic modes on the Siegel Cognitive Style Test than did the comparison groups.

Two studies in engineering education deserve some attention. Khan and D'Offyley (138) attempted to develop a battery for predicting success in engineering. They reported that prior grades, especially in physics and mathematics, made better predictors than did verbal and mathematics aptitude measures. There was no indication in the report of this study about the type of analyses which were performed to reach this conclusion.

Hoyt and Muchinsky (116) examined experiences of engineering graduates to determine what contributed to their occupational success. Ten on-the-job factors were evaluated by the graduates: 1) scientific knowledge, 2) engineering problem solving, 3) originality, 4) persuasiveness, 5) interpersonal competence, 6) managerial skills, 7) written communication, 8) oral communication, 9) precision, 10) practical judgment. Items 5, 6 and 10 were most highly rated both by graduates and their supervisors. Supervisors also stressed oral and written communication skills. Engineering courses, according to the graduates, helped give them competence on items 1, 2, 3, 9, and 10. Other courses helped for items 6, 7, and 8. Extra curricular activities were not rated as helpful for any of the 10 categories.

People differ in their divergent thinking potential as measured by the Torrance Test of Creative Thinking. Abraham (3) felt that the
quality of discussions in chemistry might be influenced by the way with differing potentials were grouped. He formed both homogeneous and homogeneous discussion groups based on the foreground coded discussions with the Aschner-Gallagher interaction and found that the groups differed. Moreover, there was no discrimination between groupings and divergent thinking potential. He found that groups were favored in homogeneous groups while high divergent thinking favor in heterogeneous groups.

Cohen (51) compared the questioning patterns of 4 science and social studies teachers. He attempted to relate the questioning patterns to critical thinking scores. As we might have expected, questioning patterns had no effect on the scores. Teachers asked more questions than did non-science teachers.

Renne et al (221) examined the impact of achievement participation during inquiry sessions. When the achievement test and I.Q. in a regression procedure were removed, it showed that achievement made no difference. Generally the higher ability students in the study gained more from the discussions than did lower ability subjects.

Nelson (197) examined the consequences of providing instruction on science concept learning of sixth graders. The non-probing treatment yielded better science proficiency scores that school differences were bigger than treatment differences. He found that the two techniques were distinguishable.

Counter intuitive events are meant to provoke unprecedented events, and students. He added to the experimental exposure a series of intuitive events and examined the effect on achievement. The main achievement gain and retention for the experimental group was significantly greater than for the control group.

Feedback on exchange of information between teacher and students can take many forms. One of them is homework. Hilligan and Kerig collected and corrected the homework of one mechanism of operation and simply collected the homework of the other. There were no significant differences between groups. The a priori hypothesis variables confounded and the conclusion reached not to assume students do not know how to use feedback may not be warranted.

In a study of the effectiveness of formative evaluation activities, Friel (79) found that when eighth grade students performed poorly, they benefited more from an analysis of the content than from additional practice items. It was important to the production of better student performance.
Halsted (100) compared the effects of student assembly of orbital models in CHEM study with the results when teachers lectured and demonstrated assembly. The active condition won. Halsted asked the students to express their feelings about which mode of instruction they preferred. Seventy-eight percent of the control group would have preferred to construct their own models rather than have the teacher demonstrate.

Studies by Murray (195) and Power (215) were also reviewed.

Piagetian-Based Studies

Using Piaget-type tasks, Kavanaugh (132) tested the validity of a proposed hierarchy through which children must pass in order to understand the concepts of speed. One-hundred and eight students in grades six, seven, and eight tried five tasks: 1) speed of circular movement with concentric circles, 2) circular path of objects fixed to a rod, 3) conservation of uniform speed, 4) speed of movement in succession, and 5) uniformly accelerated motion. He used a scalogram analysis technique to test for the existence of the proposed hierarchy. The predicted order based on Piaget’s work was 1, 2, 3, 4, and 5. The observed order was 1, 2, 4, 5, and 3. I.Q. was also related to performance. Sixty-seven percent of the sample could not complete the tasks. The stage of formal operations among this sample of students was later than the 10 to 15 years of age proposed by Piaget.

Sandra Coleman (53) worked with the other end of the age spectrum to determine whether aging produces a decline in cognitive functioning. Is there a retrogression in cognitive functioning as indicated by three tasks measuring conservation of mass, logical inference and combinatorial operations? The women in the study ranged in age from 20 to 94 years. She reported a downward trend toward pre-operational thinking with increasing age. The means for the test of logical reasoning were low for all groups, with more than 50 percent of the younger group functioning at the concrete operations level. Of course, in a study of this kind we cannot separate age effects per se from early educational differences.

Thiel (272) took groups of children who were at the concrete operation level and examined the factors that affected the way in which they made predictions. He found that children made predictions more easily using a classification rule than a seriation rule. Children in grades three through five experienced difficulty in coordinating multiple dimensions. These factors, Thiel reported, affected the child’s ability to make appropriate predictions.

Students working with Professor John Renner at the University of Oklahoma have for several years worked on the relationship between Piaget’s developmental levels and performance in science. In this tradition, Lawson (154) administered Piagetian tasks to 51 biology, 54 chemistry, and 33 physics students in an Oklahoma high school. He used scores on the tasks in a step-wise semi-partial regression procedure to predict success on examinations. He found Conservation of Volume and Separation of Variables to be the best predictors of success. Subjects categorized as concrete operational demonstrated no understanding of formal concepts. Lawson concluded that a substantial portion of
secondary school science subject matter required largely abstract or formal levels of operation but a significant portion of the students still were operating largely at the concrete operational level.

Raven (217) has developed a test for measuring logical operations as postulated by Piaget. The test appears to exhibit sound psychometric properties and could be useful in diagnosing specific patterns of reasoning difficulties. Professor Ronald Raven and his students at the State University of New York in Buffalo have worked for some time on programming Piaget’s logical operations with the ultimate goal of improving the science performance of elementary and junior high school students. Certain children may come from environments which fail to provide consistent inputs that encourage progress up the cognitive ladder. Nous (199), a student of Raven, examined the impact of a Piaget-based structured learning experience on children’s correlative thinking. Ninth-grade students did profit from the training. However, the training was of doubtful benefit for fifth and seventh graders. Hammond (101), who also worked with Raven, studied the effectiveness of using a structured set of exercises to help sixth and eighth graders learn compensatory tasks. The general problem in each task required the students to balance or make equivalent two sub-systems that involved force and distance variables. The learning program involved geometric figures, level-problems, goldfish equilibrium problems, and communicating vessels problems. Four logical operations were involved. Hammond found that on the criterion measure the experimental group did better than the controls, and that eighth graders benefited more from training than did sixth graders. He concluded that students at the formal operation stage of development can profit from this kind of training.

In support of Raven’s deficit notion, Johnson (125) showed that low socio-economic status students did not perform as well on classification tasks as did their richer counterparts.

While the Raven-sponsored studies ask whether training in Piaget’s operations facilitates science performance, other investigators ask whether science experience helps students to progress along a Piagetian hierarchy. With this latter approach in mind, G. H. Hansen (102) compared two types of elementary science programs to determine their consequence for intellectual development of sixth graders. Fifty-six sixth graders were divided into two groups. One group studied the SCIS Energy Sources unit in the first semester and the local science program in the second semester. The second group studied these two programs in the opposite sequence. All were assessed on conservation of perimeter, weight, and volume three times during the year. There were no significant differential effects that could be attributed to program except on conservation of displacement of volume, which showed up on the third round of testing in the group that had SCIS in the fall term. Hansen suggested a delayed effect of SCIS training. Presumably the second group which also had SCIS would show this effect some months later.

T. W. Brown (39) also evaluated the SCIS program and found that, in contrast to another science program, students in SCIS were superior in scientific curiosity, attitudes toward science and figural creative thinking. Hall (99) found that a group of students in England did not
attain conservation of chemical identity composition and mass after exposure to nine months of chemistry.

Phillips (212) investigated the relationship of children's ability to perform selected Piagetian tasks and their level of achievement in science as well as other subjects. There was a difference in science as well as other subjects. There was a difference in science achievement between students who performed well on the tasks and those who did not but there was no difference in mathematics achievement. Here again, although this study was under-analyzed, we can infer the presence of differential aptitudes that can influence learning.

From these studies we seem to be forced to the conclusion that students profit more from instruction in logic and abstract content only after they have passed from concrete to formal operations. Just when, on the average, they do that seems to be quite different for the various samples used in these studies. The evidence that development along a Piagetian hierarchy is stimulated by experiences in science seems to be mixed. Jerry and Mary Ayers (18) reported that exposure to the SAPA program increased the rate at which kindergarten children achieved conservation skills.

We seem to be led increasingly to the conclusion that learning to conserve one kind of quantity does not insure that transfer to new variables will happen. Conservation of mass seems unrelated to conservation of volume. F. A. Smith (253) provided yet another example with young adults who are learning fairly abstract concepts. Using programmed instruction, he taught the law of conservation of leptons and electric charge to college non-science majors to see whether this would facilitate the learning of the conservation of momentum. He did not find any significant transfer of learning from one conservation law to another, although exposure to the unit on baryons, leptons and electric charge speeded up the learning of the conservation of momentum unit.

Behavioral Objectives

Does it help, hinder or leave the student unaffected to give him behavioral objective statements prior to instruction? A review of this literature by Duchastel and Merrill 4 in the "Review of Educational Research" leads to the conclusion that there is no special or powerful advantage to this procedure. About as many studies show that behavioral objectives do not help as show that they do. If people insist on studying the function of behavioral objectives out of preference or on theoretical grounds, then they might do well to refine the research questions to be able to find out what kinds of students are helped by receiving objectives, what kinds of students are hurt by them, and what kinds of students can take them or leave them. The equivocal nature of findings reported in the publication described above as well as in earlier science education reviews should warn researchers that something is wrong with the way the problem is being conceptualized. There is

evidence in the following described studies that many students do not necessarily use the objectives or advance organizers.

The data seem to be equally equivocal among the science studies reported in 1973. Koch (142), for example, found in high school physics no difference in achievement or on a learning environment inventory between students who received objectives and those who did not. However, students who got the objectives treatment scored higher on a confidence measure. Koch said that some students used the objectives as a guide and some did not and that the disposition to use objectives did not increase on the second unit.

Similarly, Coleman and Fowler (52) found no difference in achievement with or without objectives for freshmen in a science course of non-majors. Neither the sex nor the area of matriculation nor scores on an attitude inventory changed these results. Stedman (261) reported similar results in high school biology. He found that students tested with questions geared to Bloom's taxonomy performed equally well on the first four levels of questions regardless of whether or not they had received objectives.

Kelley (134) examined the effect of three treatment conditions on performance in an audio-tutorial course in college genetics: behavioral objectives, non-behavioral objectives, no objectives. He also used the Eysenck Personality Inventory (EPI) to block subjects into five groups, so that he could by means of multiple regression determine the power of both intellective and personality batteries to predict achievement in the audio-tutorial units. Objectives seemed to help performance for some groups. He found that the group categorized as 'introvert-stables' had the highest means and less need of objectives. Here we see a somewhat more refined approach to the question of how much help objectives are to particular kinds of students.

In a less carefully designed study, Shields (245) found that behavioral objectives in audio-tutorial college biological science course did not enhance achievement. He felt the effort and time involved in writing objectives might better go into improving preparations.

In high school instruction the case for objectives seems as uncertain as it does for college instruction. Olsen (203) found in high school science that prior knowledge of objectives led to higher mean scores and better retention in both whole group and self-paced forms of instruction. Motttillio (192) analyzed the effects of presenting behavioral objectives in FSSC physics. In a series of replicate studies he used both the F statistic and t-tests with somewhat different results. Generally, there was no clear win for behavioral objectives.

Presumably prior knowledge of objectives would help students to focus their efforts and so make the learning process more efficient. However, objectives can act to impose an organization on the subject matter which is not suitable for every person. R. A. Graber (91), for example, divided 143 undergraduate chemistry students into good and poor organizers. He developed a 2500 word passage with an organizer. Some students got the organizer before reading the passage and some after. He had a non-organizer as the control group placebo. Graber
found no differences among the treatment groups. Instead he reported that skill at organizing was the key variable. Good organizers did better than poor organizers. A replicate study with 78 high school physics students produced the same kind of results. There were no significant interactions between the treatment conditions and organizing skills.

Thomas (275) found that set-induction techniques in college botany appeared to be especially helpful to females.

In a study meant to evaluate both Ausubel's organizer theory and Rothkopf's hypothesis regarding spaced questions, Feller (78) found that the advance organizer presented in oral mode only seemed to produce confusion. Among his tenth grade biology students the group which got the advance organizer and spaced-factual recall performed least well. In general, however, there were no significant differences among the experimental treatments.

Jungwirth and Tamir (130) sought to correlate teacher image as derived from earlier studies with student achievement. Students ranked 14 objectives as they thought their teachers would. The correlation between student and teacher rankings was low. Teacher Image was the best predictor. Student priorities also correlated with achievement but accounted for less variance than did three of the four areas on the Teacher Image measure.

We included the Graber, Thomas, and Feller studies in this section rather than in the section on learning to show the probable theoretical connection between objectives and organizer theory. Other reports based on Ausubel's theory appear in the section on learning. For now, one ought to recognize that most of the research on the use of objectives appears to have no theoretical base. If objectives are meant to function as advance organizers, then they appear to be doing no more good in science instruction than are most of the studies on advance organizers. These, too, have produced equivocal results.

By far the largest block of studies reviewed, audio-tutorial programs mostly seem to be of interest to teachers of college students although a few investigators used high school students as subjects. We examined these studies to see whether we could learn answers to three questions: 1) Under what conditions do self-directed, self-paced programs help and under what conditions do they seem to hinder progress? 2) What learner characteristics interact with the modes of presentation? 3) Are there any trends that would allow us to make some recommendations for practice?

If there is a single refrain sung by people who wrote and evaluated independent study and audio-tutorial programs, it concerns student procrastination. Self-starting and self-directing norms apparently are not dominant in student groups! Given this difficulty, how do audio-tutorial, individually prescribed, and/or self-study programs do in comparison to more standard programs? One way to examine the question is to list the studies as shown in Table 1 according to whether the outcomes favored the one hand the experimental treatments, or on the other hand failed to show enough difference to make the effort worthwhile.
<table>
<thead>
<tr>
<th>Investigator</th>
<th>Expt'l Course</th>
<th>Subject</th>
<th>Experimental</th>
<th>Traditional</th>
<th>No Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roth</td>
<td>Self-paced</td>
<td>College Engineering</td>
<td>+ content</td>
<td>+ content</td>
<td>fifth study</td>
</tr>
<tr>
<td>Wake</td>
<td>Audio-tutorial</td>
<td>College Electronics</td>
<td>+ content retention</td>
<td>+ attitude</td>
<td></td>
</tr>
<tr>
<td>Elliott</td>
<td>Self-paced</td>
<td>College Physics</td>
<td>+ content</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Austin &amp; Gilbert</td>
<td>Self-paced</td>
<td>College Physics</td>
<td>+ content</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnus</td>
<td>Self-directed</td>
<td>College Science</td>
<td>+ content</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brown</td>
<td>Independent Laboratory Study</td>
<td>College Biology</td>
<td>+ hypotheses formation</td>
<td></td>
<td>processes</td>
</tr>
<tr>
<td>Elliott</td>
<td>Audio-tutorial</td>
<td>College Biology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hackett &amp; Holt</td>
<td>Audio-tutorial</td>
<td>College Biology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rowsey</td>
<td>Audio-tutorial</td>
<td>College Biology</td>
<td>+ content (high achievers)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Couch</td>
<td>Guided Learning</td>
<td>Micro-Biology</td>
<td>+ content</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vander Wal</td>
<td>Audio-tutorial</td>
<td>College Biology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hunter</td>
<td>Self-paced</td>
<td>College Chemistry</td>
<td>+ content</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shinfield</td>
<td>Independent Study</td>
<td>H.S. Chemistry</td>
<td>+ content</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wright</td>
<td>Self-paced</td>
<td>College Earth Science</td>
<td>+ nature and process of science</td>
<td>+ content vocabulary attitudes</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 1** Summary of outcomes on Audio-tutorial, self-paced, independent study and programmed instruction projects. These experimental treatments were usually contrasted with traditional presentations.
We might perhaps be a little less controversial if we simply distinguished studies that showed no differences in the outcomes from those that did, without raising the question of worth.

The question of worth concerns how much difference there must be to justify the effort involved in planning and, especially, managing an audio-tutorial or equivalent type of program. Unfortunately, few of the studies dealt with this question. The question of worth can also be considered in another way. Are there particular types of students who seem especially to prosper under an audio-tutorial format of some kind? Here we should look for aptitude-treatment interaction (ATI) studies.

These are investigations which examine how learner characteristics interact with the materials or which contrast the way in which various presentation modes (e.g., verbal, spatial, etc.), interact with learner variables. ATI studies can begin to help us find out under what conditions and for what students various modes of presentation are especially useful. There are a few aptitude-treatment interaction studies included among those reviewed. Hopefully, in succeeding years the number will increase.

In connection with audio-tutorial ventures, it may be useful to distinguish between individually paced and individualized programs. An individually paced program allows students to progress at different rates. An individualized program must meet one more condition. It must make comparable content available in more than one format so that people with different aptitude configurations can find modes of presentation which are especially suited to them. Aptitude-treatment interaction studies could help us to answer the question raised in 1967 in the study by Stodolsky and Lesser: Are there benefits to be derived from matching instructional strategies to patterns of mental ability?

The need to perform aptitude-treatment interaction studies to discover what patterns of ability seem to be called upon in different subjects is illustrated in a study by Grobe. In a comparison of audio-tutorial and conventional biology programs where students were grouped in three aptitude levels, knowledge of aptitude and treatment accounted for only 9.5 percent of the variance on CEEB scores (see Table I). The 91.5 percent unexplained variance warrants some attention. The multiple linear regression procedure used in this study is an effective alternative to two-way analysis of variance since it preserves more data and makes the need for identification of other variables more obvious.

If there is a problem of student procrastination in self-paced modular programs, what can be done about it? Nelson and Bennett (198)


surmised that shorter units in a psychology course would help eliminate the problem. They set up a condition in which they organized the material to be learned in one case into twelve units and in the other, twenty-three units. Nelson and Bennett concluded that more small units result in better productivity. Students paced themselves more regularly with shorter assignments and they seemed to investigate the material in more depth. The cost to this procedure lies in the increased time spent taking tests on each unit. It is possible to offer an alternative explanation to Nelson’s and Bennett’s reported differences in performance. More frequent exams over smaller portions of subject matter necessarily lead to more rehearsals of content and more feedback on how one is doing.

Atkinson (15), arranged information in an Intermediate Science Curriculum Study (ISCS) level II program in two formats, printed and audio-tape. He grouped 72 eighth graders into three levels of reading ability and randomly assigned students from each level to a treatment. On measures of achievement, retention and time to complete the work in the ISCS’ self-paced program, all three audio-tape groups out-performed the printed instruction groups.

While Atkinson was entirely correct in using an analysis of covariance (I.Q. and STEP), in this instance he might have produced more information had he also employed a multiple regression procedure. We would then know how much of the variance on the outcome measures would be accounted for by knowledge of I.Q., reading ability, and treatment. There is also the potential here for an aptitude-treatment interaction analysis that would provide some basis for individualizing instruction. This analysis would provide a means for identifying exactly which students would prosper under each treatment.

In a step-wise regression procedure, McDuffie and Bruce (178) examined relationships between personal characteristics of students in a year long audio-tutorial biology program and achievement and attitude. In addition they used a step-wise discriminant analysis technique to predict membership in high and low performance groups. In an attempt to differentiate characteristic profiles for high and low achievers and for students with positive and negative attitudes toward the form of instruction, they administered the Guilford-Zimmerman Temperament (GZTS) Survey, Moore’s Scientific Attitude Inventory, the Nelson Biology Test, the CEEB, and an aptitude measure. Personality factors on the GZTS were predictors of achievement and attitudes toward science as well as the form of instruction. But emotional attitude toward science as well as CEEB math and verbal scores were key variables.

McDuffie and Bruce concluded from the small but significant explained variance in the criterion measures (23.2 percent) that the abilities, attitudes, and personalities of the majority of students were compatible with audio-tutorial forms of instruction. The approach employed in the study provides a means of identifying students for whom the audio-tutorial method is an inappropriate instructional alternative since the discriminant analysis indicated that low group membership had high post hoc predictability with respect to achievement. Audio-tutorial instruction provides an alternative to large group instruction. The techniques employed in the study could provide a means of categorizing students into treatments that would be especially compatible with their modes of learning.
In Individually Prescribed Instruction (IPI), Peura et al (210) found that the previous grades of engineering students who had come through more traditionally organized programs did not make a good predictor of success in IPI. They reported that 40 percent of the students had difficulty pacing themselves and 95 percent felt the effort was greater than the course warranted. Oddly enough, 84 percent reported they would choose another IPI course.

Campbell (47) examined characteristics of low, medium, and high success students taking part in a self-paced high school biology program where Learning Activity Packages (LAP's) were used as the medium of instruction. He suggested that LAP's may not be good for low ability, poor readers. Students who were successful with LAP's generally had higher I.Q.'s, read better, worked faster and had higher science achievement scores. Once again we are led to the conclusion that a particular format of instruction may be differentially successful. The task is to discover which variables predict success for each type of instructional format. In short, there need to be more aptitude-treatment interaction studies.

Two studies based on O. R. Anderson's research on the effects of structure in verbal communication provide some guidance concerning the way in which audio-tutorial and self-study packages might be constructed. What he calls "kinetic structure" is measured by identifying and counting the subject matter terms which appear as common elements in contiguous sentences. By altering the frequency and distribution of occurrence of the terms, one alters the kinetic structure of the presentation. In an audio-tutorial college biology course, Mathis (176) contrasted the effects of presenting high kinetic structure and low kinetic structure tapes to students. There were no differences in the time taken to complete the program. An analysis of variance on the post test scores also failed to show a significant difference. While there was no significant interaction between verbal ability and treatment, a covariance analysis indicated that a positive relationship existed between student achievement and degree of kinetic structure.

Butterworth (44), who was also a doctoral student of Anderson, examined the effect of levels of kinetic structure on affective learning in a freshman college biology course. He measured the effect of structure on content acquisition. Students learned more under the high kinetic structure format and felt that the high structure lessons were more helpful, but not necessarily more valuable or relevant. These two studies call attention to the need to consider the amount and form of redundancy in novel materials.

Brown (38), who was interested in hypothesis formation, assigned students to independent laboratory work or to regular laboratory instruction in freshman college biology. Object visualization and anxiety levels were the learner characteristics she identified as possibly related to hypothesis formation. Here again there would have been an excellent opportunity to perform an aptitude-treatment interaction.

analysis but this was not done. The results, while interesting, are not as discriminating as they might have been. Independent study did have a direct effect on both inductive and deductive hypothesis formation. High anxiety students in general did not perform as well as did low anxiety students.

McElhattan (179) developed and evaluated the effectiveness of an auto-instructional chemistry course for prospective elementary teachers. All the students exceeded the minimum criterion score on the content test. Students also improved their performance on the Test of Understanding Science (TOUS) and the Processes of Scientific Inquiry Inventory.

In a conceptually more interesting investigation that employed principal component analysis, Barry (24) contrasted student performance on a traditional presentation of topics on sound waves with performance on a linear program of written material arranged in a knowledge hierarchy as described by Gagne. In three first-form classes he obtained test scores on the topic, mid-year science scores, number of errors in working the program and I.Q. He found the error rate was high and that many students did not correct their errors. He noted that the error rate did not seem to be significantly related to the aptitude measure, a fact which he felt calls for further research. He noted, however, that higher error rates were associated with lower scores.

Burney (43) developed and evaluated programmed materials for college physics and engineering students who were learning about amplifiers. This dissertation contains a review of programming methods and relevant learning theory, but the analysis of data leaves much to be desired.

In a study meant to contrast individualized science instruction via ISCS with traditional science, Heffernan (104) unfortunately confounded content and method variables with the result that his conclusions concerning the efficacy of individualized instruction are largely unwarranted. What he in fact did was to race two curricula, ISCS and the New York State Science curriculum, against each other. He used the TOUS, the Watson-Glaser Critical Thinking Appraisal and the Student Science Attitude Survey. Moreover, he reported that the teacher of the ISCS may have influenced the New York State program teacher to increase the amount of group activity. Heffernan might have recovered valuable information despite the design problems he had, had he performed a multiple regression analysis to find out what part of total variance was contributed by each of the factors: curriculum type and pacing.

The ISCS science program for grades seven through nine requires students to direct their own progress for considerable blocks of time. E. S. Stallings (260) used the Tab Science Test to compare the achievement of ISCS students at each grade level with that of students in what he calls a traditional program. He found no differences at the seventh and eighth grade levels but did find a difference in favor of non-ISCS group at the ninth grade. The Tab Test presents an interesting approach to evaluation but, as Stallings points out, its reliability is rather poor and one must be guarded in interpreting results based on it. If, however, the results do reflect a real difference, then we may have here another instance where students do not direct their own learning very well.
Dunstan and Johnstone (68) found that 14 to 15 year-olds had difficulty learning chemistry which was presented in a programmed instruction format.

Smiley (251) used the audio-tutorial method to give secondary biology students freedom to choose whether to come to class. He found that students who were not compelled to come to class performed as well as did those for whom attendance was compulsory.

Couch (58), on the other hand, reported that in a microbiology course where students had what he called guided learning plus lab, performance was not as high as with the traditional lecture plus lab. Based on his data, Couch concluded that students want to be spoon fed and do not like it when they are not.

In a study of the Individually Prescribed Instruction (IPI) system, Rookey (226) found no differences at the early grades on a variety of attitude and performance measures. He noted, however, that middle-level students had higher scores on measures of creative tendency, self-concept and attitude toward school than did the controls. For younger students, the controls had better attitude scores but the IPI students had better self-concept scores.

Shaver (242) found that students in an earth science course who learned through independent study performed as well as did those who received standard instruction. He observed, however, that in the following year the academic achievement of the independent study group was significantly higher than for the controls.

To a group of engineering students who had taken the Myers-Briggs Type Indicator, Smith, et al. (252) presented a self-paced program. They sought to answer a question more researchers should consider: namely, for what kinds of students are certain treatments helpful? They produced some interesting findings. The MBTI has a two-way classification scheme, Extravert-Introvert, and four pairs of preferences: sensing (S) - intuition (N), thinking (T) - feeling (F), and judging (J) - perceptive (P). Seventy-three percent of the students completed the course in thermodynamics in 11 weeks. The N-types liked the self-paced formats while the S-types preferred group work. T-types preferred lecture while F-types chose the self-paced format. Introverts had higher GPA's. F-types took smaller course loads and attended more help sessions. T-types took heavier course loads. No significant correlations were found between the Extraversion-Introversion variable and scores which indicated the students' preference for various teaching methods. Smith, et al. concluded that the composite type for whom self-paced formats are desirable is NFP (intuitive, feeling, perceptive).

Kelley (134) administered the Eysenck Personality Inventory (EPI), among other measures, to find out what personality variables would predict achievement in audio-tutorial units on genetics. He found the extroversion-introversion dimension of the EPI contributed significantly to total variance. Introvert-stables had the highest means. The best predictors were grades from a prerequisite course and SAT math scores.
Wright (303) compared performance of college students in individualized and standard earth science courses for non-majors. He assessed changes in understanding of the nature and processes of science, acquisition of earth science knowledge, and attitudes toward science. He found no differences in attitudes. The traditional group acquired more earth science knowledge and vocabulary. The individual learning group gained more understanding of the nature and processes of science.

Hogg (112) did a survey to determine attitudes toward programmed learning in college chemistry and found, as did some of the investigators previously cited, that, on the whole, students liked self-paced instruction. It should be noted, however, that ATI studies indicate this mode of instruction may be inappropriate for some students. Generally the self-directed or self-propelling requirements of self-paced ventures seem to present problems for many students who find procrastination to be the biggest obstacle.

In an aptitude-treatment interaction study (ATTI), Ann Lee (157) attempted to distinguish behavior traits of students who learned best under CAI from traits of those students who learned best in a traditional laboratory physical geology course. In addition to biographical data, she administered the California Psychological Inventory (CPI), seven short-term achievement tests, a long-term achievement test, and an attitude toward CAI scale which she correlated with CPI traits and with performance. Johnson-Neyman tests of the two ATI hypotheses showed no regions of significance between the traits and the two treatments. Generally, CAI in geology appeared to facilitate learning for males, science-related majors, and students with previous computer experience. Females, liberal arts students, and those with no previous computer experience did better in the traditional format.

Thomas Smith (256) sought to identify relationships between performance and selection patterns of students learning chemical analysis through an on-line computer delivery system; selection patterns refer to positive and negative instances of the concept. He identified two patterns, attribute centered and instance centered. In the former, attribute centered, the learner selected the results of one test solution across all instances before moving to a second test solution. In the instance-centered strategy, the student selected the results for all test solutions on one instance before moving to the next instance. Smith varied the task characteristics by presenting the information in linear and matrix forms. He also presented feedback in these two forms and he altered the complexity of the task by withholding a number of relevant attributes. Each student chose what information he wanted. Smith reported that while there were considerable between-person differences, each learner tended to maintain a particular selection pattern across similar tasks. In this study, which has considerable theoretic impact, Smith categorized the observed selection patterns into four groups: 1) analytic-attribute centered, 2) analytic-instance centered, 3) global-attribute centered, and 4) global-instance centered.

Summerlin and Gardner (265) compared achievement in two weeks of chemistry under CAI and non-CAI regimes. The GAI group achieved higher scores on the content measure. Cunningham and Fuller (59), in a less well designed comparison of performance in a CAI physics course and in
standard physics, found no difference between the two groups. They concluded that preparation of CAI involved too much work for very little gain.

Wolff (299) developed and evaluated a new CAI language, called "CHIMP," to determine whether it would function for interactive tutorial and drill lessons as well as be suitable for the study of learning.

The potential of CAI for examination of science learning processes probably should receive more attention in science education research than it has to date. It makes possible the study of decision processes.

Formats for Instruction

A monolithic conception of intelligence or aptitude is rapidly disappearing. In its place are sets of special aptitudes: verbal, visual, spatialization, numerical, and a host of others. Each branch of science and even sub-disciplines within a field may draw heavily on specific combinations of aptitudes. Ann Roe noted that within the field of physics, the prominent scientists she interviewed differed markedly in their techniques for working out ideas. Some used diagrams and schematics to help them think, others never did. Similarly, students may exhibit a great variety of preferred modes for receiving and processing information. The kinds of modes in which they are required to return the results of their thinking to the teacher usually are verbal and numerical. In any case, it may be desirable, now that an extensive pool of curriculum materials in science exists, to pay more attention to the variety of ways these materials could be presented to different kinds of learners.

First we consider different modes of presentation (e.g., oral and visual) and ask whether the investigators concerned tried to distinguish students who benefited from a technique from those who found it inhibiting or at least of no help in learning.

If eighth graders do not read very well, will presentation of information on audiotapes help them to perform more successfully in science? Atkinson (15) showed that in the ISCS program when students who did not read well received instructions orally, they performed better than did their poor reader counterparts who did not get much help. With a sample of first grade children, Hibbard (106) attempted to remove poor reading ability as a factor in learning science concepts by supplying 18 audio-tutorial lessons. These lessons included work-sheets and manipulative materials. Children in his sample did develop concepts without benefit of reading in the usual sense.

However, Moore (187) found that biology students who were poor readers were not especially helped by audiotapes that supplemented their texts. Neither achievement nor attitude scores differed from those of the control students.

---

Hibbard (105) also showed in a second study of first graders that a combination of pictures and actual manipulation produced the most learning if these techniques were coupled with discussion. In both studies he reported that the correlation between reading ability and science learning was not significant. As others have done before him, Hibbard suggested that science experiences allow the non-reader to build verbal facility. In this connection he found a strong positive correlation between scores on a science test and verbal predictive ability.

What do we gain by giving students materials to work with in a laboratory type exploration as compared with sending them to the library to gather the same information? Merkle (184) reported that inservice teachers who had a laboratory experience with pendulums gained more at the end of six weeks than did those who studied pendulums by reading about them.

What effects do pictures have on learning? Here we need to consider whether the pictures are just decorative, whether they are redundant with text or whether they truly substitute for words in text. In short, how much of the message does the writer mean to have transmitted through pictures? Holliday (113, 114) has reviewed the research on use of pictures in science education. For people interested in this topic, his review is helpful.

Particular attention should be given to those studies that examine the relation between the learner's characteristics and his ability to make use of pictures. If the picture is meant to carry the message, then the artist and the authors work together from the start on production of curriculum materials. Each shapes the thinking of the others. The ISIS project, a new NSF funded program directed by Ernest Burkman at Florida State University in Tallahassee, has proceeded in this way.

Holliday reported that pictures which are adjunct to the text do not always help. If they are simple, on the order of diagrams, there may be a facilitating effect. He warned that the influence of pictures is complex, but at least there should be some correspondence between the kind of concept to be presented and the picture. Some concepts do not lend themselves well to pictorial presentation while others probably ought to be presented with graphics.

Pictures may be used not only during instruction, but also their potential use in evaluation warrants attention. Reese (219) constructed a pictorial test for Blue Version BSCS biology. He supplied validity and reliability data. The test was both individually and group administered with similar results. It would now be interesting to construct a verbal equivalent of this test and to find out whether students with different verbal and spatial aptitude patterns perform in the same way on both forms of the test.

Voelker (291) suggested that the kinds of pictures which help may differ depending on whether students come from urban or rural settings. He collected feedback from several groups of children concerning which uses of particular pictures helped. Probably feedback from students in other than a typical test mode might be helpful in the whole area of pictorial communication.
Talley (270) found that students in freshman chemistry did more critical reasoning when they used molecular model kits to help them understand molecular structure. Ian Thomas (274) found that films helped Australian secondary school students learn Coulomb's Law.

B. W. Hill (108) found that audio-visual slide tape units helped students in general college chemistry to learn chemistry. In one treatment condition, students were free to use or not use the tapes as they pleased. Since this treatment group did as well as the group which was required to use the A-V units, she concluded that students were good judges of their needs. They used such supplementary resources if they felt they needed them.

Three studies on the ways in which planetariums help students learn astronomy and aerospace concepts are reported in this section. Akey (6) found in a one group pre-post design that second graders understood 24 of 39 concepts prior to exposure to the planetarium. There was a positive correlation between time spent post-teaching planetarium concepts and retention. Sunal (268) also used second grade subjects and found significant gains in all goal areas. However, a group exposed to an astronomy unit, as compared with an astronomy-planetarium experience, did as well. He concluded that the planetarium functioned largely as a remedial and review device. Reed (218) also found that the planetarium as a demonstration chamber accomplished little with college students. When the planetarium was used in conjunction with classroom chalkboard drawings and celestial globes, more learning occurred.

We could find no work to date on the use of pictures, cartoons, etc., to convey any of the affective and valuing aspects of science.

Lerch (159) presented 129 college physics students with a divergent lab and 35 controls with what he called a traditional presentation which was largely deductive. He used a questionnaire to tap six factors: 1) qualitative and quantitative measurement, 2) choice of experiments, 3) interest in experiments, 4) self-sufficiency in lab, 5) verification, and 6) use of models. The divergent treatment group responded differently from the controls on factors 2, 3, 4, and 5.

In a high school biology course, Egelston (70) found that groups exposed to inductive instruction did poorly at first but finally surpassed the deductive group in achievement. The controls, however, exhibited a better attitude. As far as the rating of the quality of the learning environment went, the controls scored better on intimacy, satisfaction, and diversity indicators while apathy, disorganization, goal direction and formality were factors rated strongly by the inductive group.

Linz (161) contrasted inductive and deductive chemistry approaches. The inductive approach went from specifics to generalities. The deductive mode progressed from general to specific. The deductive mode produced generally better results.
In this section we include studies that examine outcomes of instruction in various content areas. Outcomes can be cognitive or affective or both. Those papers which reported standard kinds of achievement results are discussed in the first part of this section. More investigators are taking an interest in attitudes, preferences and values that develop or are influenced by instruction. Studies where such variables receive attention appear in the second part of this section.

From this latter group we learn that the pass-fail option does not seem to facilitate learning or to particularly influence student liking for science. Dogmatic people find modern science programs with their emphasis on inquiry processes difficult. There is evidence to suggest that self-concept and personality changes rarely occur from exposure to science of the durations typical in these studies. However, there is evidence that students who learn an analytic mode of inquiry in junior high school still retain the mode when they are seniors. There is also an indication that science is a great equalizer with respect to people who differ in self-concept. People with poor self-concepts do as well as those with more positive views of themselves.

In an interesting investigation, Bullock (42) compared the relative effectiveness of three different types of elementary science programs: Science - A Process Approach (SAPA); Environmental Studies Project (ES); and the Laidlaw textbook series. Twenty-seven teachers and 512 sixth graders in a southern school system took part in the projects. The TAB Science Puzzler Forms B and C were administered as pre- and post-tests. On selected problem solving skills there were no differences between SAPA and Laidlaw groups but both these groups out-performed the ES group. Both SAPA and Laidlaw programs produced improvement in problem solving skills. There is no report as to whether the investigator verified how much science was taught in each condition or whether the processes used reflected the intents of each program.

Renner, et al. (223) did a series of investigations of the Science Curriculum Improvement Study (SCIS), an activity-based elementary science program. The paper, which appeared in School Science and Mathematics, is a composite of dissertation studies by Renner's students. The general tenor of the studies favors SCIS as a program which encourages development of science processes and language. Development along a Piagetian continuum may be encouraged by participation in the SCIS program. (See the section on developmental studies for further discussion of this point.)

Daug (60) compared sixth graders' performance on an earth's atmosphere unit where the ostensible difference in treatment was simply whether or not a class used one text or many texts. With the use of the Cloze procedure to measure gains, Daug found no performance differences between the two groups. However, teachers found the management with five versions of a program rather difficult. In short, the multi-text treatment did not help the subjects in Daug's study any more than a single text did. If the range of reading skill in the group were not large, then one could expect little advantage from a multi-text approach.
How does the duration of training relate to retention? In a well conceptualized and analyzed investigation, Audrey Tomera (279) examined the retention and lateral transfer of the processes of observation and comparison. She asked whether these processes once acquired would be retained over intervening non-training periods of three months to a year. Data were collected from a rural junior high school and from an urban junior high school. Achievement was measured by the Scientific Observation and Comparison Skill Test - Twig or Algae Forms developed by H. R. Hungerford of Southern Illinois University. With a multilinear regression procedure for data analysis, Tomera showed that skills are well retained at the end of a year. The duration of this retention did not seem to be a function of the length of the training period. With these results it is not surprising that spaced review, a treatment which she tried, provided no additional gain in retention.

In connection with the Tomera study it is of interest to note that Knapp (140), who also worked with Hungerford, examined the transfer of classification skills by prospective elementary teachers. He asked whether training on classification with one category of materials facilitated learning to classify a different set of materials. Knapp provided training with twigs and found a significant transfer of learning when students were tested on the algae form of Hungerford's test. Students learned to classify algae more rapidly as a result of the twig experience.

L. Allen (9, 10) found that five percent of the variance in achievement between SCIS and non-SCIS students could be attributed to the SCIS program and could two percent of the motivation variance. This study which appeared in the Journal of Research in Science Teaching, Vol. 10, No. 2, also came out with a different title, a shift in table format in Science Education, Vol. 57, No. 2. Essentially Table VII in JORST is a condensation of Tables V through VIII in Science Education.

Scott (237) became interested in whether students who had had several years of exposure to a version of Suchman's inquiry training program in elementary and junior high school developed and maintained a more analytic style than did a matched sample who had had science but had not had this training. He found 26 high school seniors who had experienced the training in elementary and junior high school. They did exhibit more analytic behaviors.

What is especially desirable in the analytic mode? In contrast to global modes of thinking, the analytical style apparently contributes to overall achievement and may be especially helpful in chemistry. Scott's study suggested that students can be taught to acquire a particular style of inquiry. What remains to be determined is how that style influences the way students function in different disciplines.

One study may be worth mentioning in the context of students with learning problems. Gilbert (86) applied four different readability formulae to basic elementary science text books: the Spache, Dale-Chall, Lorge, and Fry. Each gave slightly different results, but generally Gilbert reported that the tendency was for science books to be slightly more difficult in reading level than they ought to be according to the formulae. We must be aware, of course, that the
Special terminology of science tends to inflate these formula results. The readability formulae were developed for general rather than for technical reading. Nevertheless, the warning is there for people who prepare instructional materials in science.

Grosmark (93) sought to determine how more laboratory work in high school chemistry influenced achievement, laboratory skills and attitudes. The additional experiments were in some respects redundant and students performed them during free time. There were no differences in attitudes or achievement but there was a difference in laboratory skills in favor of the group doing more experiments.

Vanderschmidt (288) demonstrated that eighth and eleventh grade students can learn to criterion level the cardio-pulmonary resuscitation curriculum. They needed more mannikin practice than they normally get in that program and also mnemonic devices in the practice protocol for compressions to establish rhythm.

Casas and Fry (48) showed that 14-year olds who participated in a critical thinking program which emphasized experimental design and evaluation of inferences did better than a matched group which had no training. Students who were in a pre-formal operations stage of development learned to design experiments and to criticize designs. This design which puts a training condition against no training is rather unproductive. A pre-post group design where each subject is a control on himself would yield more information.

In zoology, Morrison (190) asked which of three instructional sequences would produce more learning and found no differences. Here we infer again that; within rather wide margins, the learner imposes his own organization on the content to be learned.

In a comprehensive analysis of the Nuffield science courses to determine what mathematical competencies they require, Malpas (171) found the following mathematics topics to be important: ratios, rates, proportions, graphs and formulae; measurement and statistical treatment of data; and some geometry and trigonometry. He surmised that some of the topics were introduced before students had achieved a Piagetian developmental level that would allow them to grasp the concepts. So here, as in studies described in the section on developmental aspects of learning, we find evidence that concepts may be being presented at a level of abstraction which is inappropriate to the developmental stage of the majority of students.

R. Denny (62) described an attempt of teachers from an inner city school to improve performance in chemistry by concentrating on basic mathematical skills needed for chemistry. The study illustrated that trade-offs are involved. On the ACS-NSTA High School Chemistry Tests the students showed improved performance on questions that required basic mathematical skills, but this happened at the apparent expense of items relating to basic chemistry concepts and skills.

Bedderman (33) examined the relative impact of two motivational sources on fifth and sixth graders' problem solving behavior in which the tasks required them to learn to control variables. He compared
cognitive conflict condition with an external reinforcement condition. His three conditions consisted of: 1) training plus records, 2) training plus cognitive conflict, 3) no training. The purpose of the training was to teach the 31 fifth and sixth graders who had a mean age of 11.8 years to control variables. Piaget contends that the ability to control variables must wait until combinations; reversibility, scoring, etc., are in the individual's repertoire. Typically this means controlling variables would normally not be well done until a person has reached the formal operations stage of development. Half of Bredderman's group could not learn to control variables. He found no difference in performance outcomes between the cognitive conflict and external reinforcement conditions. He suggested, as Rowe did (231:Chapter 10), that external reinforcement may not be necessary in elementary science. It may even block certain desired outcomes.

With concern for elementary students' feelings about the relevance of science to them as the motivation for her study, Feerst (77) presented the COPES program in two ways to 12 intact classes. One group received constant encouragement to look for applications and examples of things studied in situations outside the classroom. The other group had no special stress given to the relevance of what it was studying conditions outside the classroom. After five months, attitudes of both groups toward the relevance of science were the same. The experimental group, however, showed better retention.

Fisher (80) compared performance and attitudes of junior high students in a traditional science class with those of students using the Rand McNally curriculum, Interaction of Man and the Biosphere (IMB). With use of analysis of covariance procedure to compensate for non-random assignment to classes, Fisher found that IMB won on achievement and reading and produced slightly better attitudes.

In a study of affective aspects of the ISCS Level I program, Lauridson (152) used a pre-post design with a comparison group to examine attitudes toward science, self-reliance, ranking of science compared with other subjects and classroom environment. Since the entering states of these variables may influence the final state after treatment and because they are probably not independent, Lauridson used a discriminant analysis technique. We found that both the ISCS and non-ISCS groups ranked science lower at the end of treatment. Both groups improved somewhat in self-reliance. The non-ISCS group became more negative about science as a vocation. The-ISCS group became more positive about the nature of science. What we cannot infer from the design of Lauridson's study is whether the changes reported are largely maturational. Is the shift in self-reliance generally characteristic of this grade level anyway?

One goal of science instruction seems to be related to improving the understandings which students have about the scientific enterprise, the way in which scientists work, and what is unique about science as a socio-economic enterprise. Reis (220) developed two instructional procedures meant to nurture the growth of this goal. For Physical Science I students at Stanford he prepared two programs. In the first, students functioned as a team of scientists to develop hypotheses and perform experiments. The second program involved the students in visits
and discussions with scientists and in the study of scientists' work. The two programs were equally effective in attaining the objectives. Hall (98) gave one group of education students 11 case histories of scientists to read while the second group got a science text to read. The experimental group achieved higher scores on the Test of Aspects of Scientific Thinking (TAST) and on the sub-test which measured ability to draw conclusions.

Vermaelen (290) found that participation in a high school science course on a pass-fail basis was not an effective procedure for producing more positive attitudes concerning self, teacher, learning, and the academic subject. Student attitudes toward the grading option tended to reflect the teachers' attitudes. Gatta (83) compared performance in eight high school chemistry classes where four groups operated on the pass-fail option and four received grades. Pass-fail groups did less well. There was an interaction between ability and treatment. High ability students achieved more and had better attitudes in the conventional system. There was no effect of treatment on middle and low ability groups on achievement or attitudes.

The Egelstons (71) were interested in whether junior high school students in general science and earth science would accurately predict their own performance on tests. Ostensibly, self-evaluation could lead to changes in course involvement and study procedures. They found that self-appraisal methods can be learned. Generally, higher ability students made better predictions. This investigation raises some potentially interesting questions concerning the perceptions and reward histories of middle and poor performing students.

Himaya (109) found that students who scored at the close-minded end of a dogmatism scale did not gain as much understanding of science and scientists (TOUS test) as did their open-minded counterparts.

Mullins and Perkins (193) exposed 34 students to an innovative 16-unit college biology program which placed emphasis on application of biology to social problems and on emotional as well as intellectual growth. On the Personal Orientation Inventory, a measure of self-actualization, there were significant positive changes on 9 of 12 scales. Greatest changes were in the shift to more inner directedness and to acceptance of one's own aggressiveness.

In a similar vein, K. Kuhn (151) used the Theoretical Orientation (TO) scale of the Omnibus Personality Inventory and found that unless instructors emphasize attitude changes, none will occur. Exposure to biology and psychology courses did not change TO scale means.

Schock (235) has suggested, as a result of a study of 28 high school biology classes, that there may in fact be a possible inverse relation between cognitive and affective scores. Waldstein (293) tested the proposition that male high school students who exhibited a large discrepancy between their self-concept and their ideal self would be less able to apply principles of dynamics than would their relatively less afflicted counterparts. Research by Ann Roe and
B. T. Eiduson on profiles of physical scientists provided the basis for hypothesizing that the discrepancy score would be positively correlated with achievement in dynamics. The hypothesis was not confirmed. The ability to apply principles of dynamics was not a function of the gap between one's ideal self and current self-concept. Apparently the debilitating anxiety and rigidity which often hampers performance of people with poor self-concepts was not operative here.

Sears (238) measured the affective changes in students who took part in an innovative engineering course. He measured changes in (1) student's perception of himself, (2) debilitating and facilitating anxiety, (3) internally-oriented characteristics, (4) achievement motivation, and (5) understanding of engineering function. For instruments he used Osgood's Semantic Differential, the Rotter Internal-External Locus of Control Scale, Edwards Personal Preference Schedule. He compared responses of the junior students in the sample with those of a set of practicing engineers. The experimental group moved toward a more internal locus of control.

Gibes (94) also sought to examine the affective outcomes on students of a course in college genetics in which the general aim was to improve scientific literacy. Students rated the learning environment as more important than the course content.

In an Environmental Action Program (EAP) at the University of Iowa, Sheldon (244) found that students' self-concept scores improved. Participation in the EAP yielded some changes on a social issues attitude survey.

Other evaluation and curriculum studies reviewed were by Auckland and Weatherup (16), Bruschwig (40), DeBlanc (61), Joels (124), Krause (148), Loc. (162), Miller (183), Ryder (233), Snyder (257), and Yost (307).

Tests

Years ago, shortly after PSSC physics appeared on the American scene, Robert Heath, who was with Educational Testing Service at the time, developed a pilot version of what he called a cognitive preference test. Some students prefer abstract generalizations, Heath argued, while others prefer fact or knowledge types of statements. Programs may differ in the development of preference for generalizations as opposed to facts. At the time Heath constructed a number of physics statements or problems and provided a list of possible responses. Students were to choose from each set of responses the one they preferred. All statements concerning each situation were true but the statements varied in their level of generality. He thought that one

---


way in which PSSC, and indeed what were then new science programs, differed from their predecessors lay in their emphasis on relationships and principles rather than on facts. A cognitive preference test would disclose whether, in fact, such a difference existed. Unfortunately this line of work was not pursued beyond the pilot venture until Ronald L. Marks took it up some years later.

In this review we are glad to report one more venture in the measurement of cognitive preference. Mandelare (172) constructed and examined the psychometric properties of a general science cognitive preference instrument. He also constructed an alternative form of the Marks Cognitive Preference Test for high school chemistry. His objective was to produce a test that was easier to administer and that would be less time consuming. In addition, Mandelare did what we wish more of the others whose work is described in this section had done, he performed a factor analysis to determine the construct validity for the four factors incorporated in the test: memory (M), application (A), principle (P), and critical questioning (C). Since various aptitudes may be a factor in the preference pattern, Mandelare also correlated, among other aptitudes, verbal reasoning (VR), numerical ability (NA), abstract conceptions (AR), with the preference measures and the New York State Examination in chemistry. No clear principle (P) factor emerged but the first factor was significantly correlated with science interest. We have given considerable attention to this study because it seems a productive line of inquiry and Mandelare has provided good guidance for people who are interested in the construct validity of instruments that measure attitudes and preference.

Badaracco (20) developed an interesting test procedure intended to find student preference for various aspects of outdoor environmental education. He used 12 cards which depicted outdoor environments. These were shuffled and students then placed them in order of preference. He used this modified sort technique with 328 students in grades 1-12 to get comparative ratings. Mammals, birds, fish, Indians and geology were the top five topics, in that order.

Seymour and Bingman (240) also sought to develop a "views and preferences" measure for use with tenth grade biology classes. This instrument would measure the degree of implementation of an inquiry approach. Here the intent was to distinguish preference for the inquiry role approach (IRA) and a non-IRA state of nature. Student attitudes toward social interactions, cognitive operations, and teaching procedures were taped in a Views and Preference Measure. Seven hundred IRA and 520 non-IRA students (BSCS biology, standard text-laboratory approach) took part in the validation study. If others use this instrument, we recommend that they perform a factor analysis to determine whether the inquiry role variables which it purports to measure do separate out into identifiable clusters or factors. When we suggest factor analysis, we are implicitly raising questions concerning the construct validity of the test which ought to be pursued.

Skinner and Barcikowski (250) were obviously aware of such questions when they made a revision of the Reed Activities Checklist which they administered to seventh and eighth graders. The Reed instrument measures interest in biology, earth science, and physical science. In a factor analytic approach, they identified the principal components, performed a varimax rotation to arrive at a set of five factors which they labeled as follows: 1) passive earth science, 2) passive and active biology (woody and birding), 3) active physical science (tinkering), 4) active earth sciences, and 5) passive physical science. This revised instrument, if it continues to show a clear factor structure, could then be used in a multiple regression procedure in studies which try to predict success in general science.

Eugenia Koos (144) reported another in a series of tests developed by the Mid-continent Regional Educational Laboratory (McREL) to measure inquiry skills. This instrument was developed to measure outcomes of TOPIC 7 of Explorations in Biology (EIB). Psychometric properties of the test which are generally sound are given in the appendices.

Leopold Klopfer (139) presented an interesting paper at the 1973 meeting of the National Association for Research in Science Teaching in which he described a conceptual framework for constructing measures of the affective aspects of science. He used a phenomena-behaviors grid as a basis for generating items.

Aikenhead (5) described a test construction procedure which he feels may be useful for people interested in formative evaluation. He gave the TOUS and the Science Process Inventory (SPI) to 921 students in Project Physics before and at the conclusion of instruction. He selected items that showed a significant shift from pre to post-test and combined them into a single test. Then Aikenhead examined some of the psychometric properties of the new instrument by giving the test to 64 students. The new test had greater predictive validity for Project Physics than did either TOUS or SPI. We would recommend an additional step. The new instrument should be subjected to factor analysis to determine its underlying structure. Subscores on the factors which might emerge could then be used in a prediction of success procedure.

Steiner (262) followed a sound procedure in developing a questionnaire meant to tag some science values held by high school seniors. He used a factor analytic approach from which emerged seven factors: 1) regard for human life, 2) pessimism, 3) cooperation with nature, 4) concern for population, 5) personal responsibility for societal problems, 6) optimism, and 7) individual freedom. He estimated reliability with data from two samples. In general he found that the more science people had taken, the less neutral they were on each of the factors.

Sister Agnes Sun (267) developed a multi-phased set of procedures to get at student attitudes in a physical science course for preservice teachers and to relate these to the degree of implementation of a particular kind of instruction. While the totality of the conception is to be praised (especially her attempt to relate variables to each other), the tiny N of nine students in each of two treatment conditions means that we can conclude nothing about the value of the procedures at this point.
Two investigators produced instruments focusing on elementary school science. Toth (281) felt people need help in determining which elementary science program is most likely to meet their needs and goals. With this in mind he developed a nine category self-assessment instrument. As yet it has had little validation. Kaur (131) added another instrument to those reported in past years which attempt to measure process skills of elementary children. As in most of the other instruments, observation and classification were the processes assessed. The two instruments are meant for use with grades one through three. Final versions of the test were administered individually. Reliability of the observation measure was reported to range from .86 for grade one to .94 for grade three. Reliabilities on the classification instrument ranged from .59 for grade one to .62 for grade three. Correlation between the two measures was .86. With this high a correlation between observation and classification, how much do we know about classification? We think it is reasonable to suggest that Kaur partial out the variance on the classification measure which is due to observation and then examine what is left. That remainder, if any, may throw some light on classification as a process in and of itself.

In chemistry, Coley (54) reported that grades in a chemistry prerequisite course were better predictors of success in junior college general chemistry than was the Toledo Chemistry Placement Examination. Penna (208) correlated I.Q., College Entrance Examination in Chemistry (CEEB), and the ACS-NSTA Forms, 1967 and 1969, for 30 students. For this small sample of students he found the correlation between these two national tests to be high.

Stokes (263) found in college chemistry that students with lower aptitude ultimately performed better on the ACS-NSTA, Form 1967, if they had short frequent tests. It made no difference to higher ability students whether the tests were spaced over longer periods. The total number of items responded to under either condition was the same, only the timing of administration differed. She reported that low ability students were more likely, in general, to make larger gains in chemistry achievement, regardless of test spacing. It is possible that what she was reporting was a simple regression effect. One interesting facet of this study concerns the relation between test spacing and student facilitating and debilitating anxiety as measured by the Alpert and Haber Achievement Scale. Students tested under the conventional pattern (longer intervals between tests) demonstrated more situation specific maladaptive anxiety. Drive for achievement was higher among students tested more frequently.

Tilford (276) thought that Negro students in a black college would have different attitudes toward science than would white students so he developed a 35-item instrument to assess attitudes of blacks. Except for items relating to the role of Negroes in science, whites and blacks responded in the same way.

Schofield (236) reported that teachers in England encourage students to guess on tests and discussed the consequences of this advice for the interpretation of test results. Selmes (239) reported on the influence A-level biology has on attitudes to science. He contrasted Nuffield and non-Nuffield A-level courses.
Maddock (167) developed and tested an instrument to measure attitudes of Papua New Guineans towards investigation, control, and manipulation of natural phenomena. Versions were prepared in English, New Guinea Pidgin, Enga and Hiri Notu' languages. Reliability characteristics were satisfactory. Maddock found that formal schooling was creating an attitude gap between the young and the old. Science teaching probably contributed to more positive attitudes toward investigation, control, and manipulation than was the norm among the elders. The procedures for constructing the various versions of the instrument might be instructive for investigators interested in multicultural analysis.

Physics

Why should physics enrollments over the past five years be on a downward swing? This problem has bothered many people and provoked numerous speculative articles, but not very much systematic investigation that has shed any light on the subject. Perhaps the most interesting possibility appears in a study by Bridgham (35) who wondered whether grading practices of physics teachers differed from grading practices of other subject matter high school teachers. He found that a great disparity existed between physics and chemistry teachers' grades and those of teachers of other subject matters. He suggested that, within limits, enrollment and grading practices may be connected. The disposition of physics and other science teachers to be hard graders, even more so than mathematics teachers, warrants some examination.

Dietrich (66) also examined grading policies by comparing public schools having high enrollments in physics with those showing low enrollments. He obtained samples of students in two such schools and grouped them into four categories according to how their physics grades compared with their science grade point average and their overall grade point average. He confirmed Bridgham's report that physics grading was severe but could not attribute the difference in enrollment in the two sampled schools to grading practices.

In a general survey of the status of high school physics, Ivany, et al. (120) found that, while enrollments in chemistry were beginning to climb, such was not the case for physics. These investigators also took one week teaching samples for four categories of physics teachers, those doing PSSC, Harvard Project Physics, something they called modern-traditional, and traditional in combination with pieces of new curricula. We may infer something about teaching style that could be related to who enrolls in physics and why. Virtually all the teachers in their study, except those doing Harvard Project Physics, had high direct modes of teaching. They also found that the percentage of non-college bound students in physics classes was extremely low. In short, physics is not being made attractive to many students. In the course of the study they interviewed teachers, counselors, and students. Students actually taking the courses had favorable attitudes toward science, and teacher background was better than expected.

One indication of the high school impact on decisions made by students when they reach college can be gleaned from a study by Brown and Elliott (37). They found that most of the 1,108 students in their
college physics sample took physics because it was required and not because they were interested in it. Generally, students who had more favorable responses toward their college physics courses were those who had taken physics in high school. Female representation in physics courses is ridiculously low. They make the remark (37:54) that if more social, political and historical areas of physics were treated perhaps more females would elect physics. However, they fail to indicate whether their suggestion simply represents their conception of what women would like to study or has some basis in the data of their survey.

Pharris (211) surveyed senior high school students in nine Connecticut high schools and found that most of those in physics classes did not elect it, they were scheduled into it.

It would appear then, that membership in physics classes results largely from response to a draft.

Aikenhead (4) has gone off on a somewhat different tack. He examined the ramifications of good qualitative feedback. He used the McNemar chi square item analysis procedure to identify areas of knowledge and specific ideas for which students showed a significant increase or decrease in understanding over a period of instruction. Thus it was possible to develop feedback specific to a project, a teacher, a student. Project Physics students, in contrast to other physics students in his study, appeared to exhibit better performance in values of science, science-society interaction, science tactics, and function of science.

If certain courses do attain a reputation such that the students come to value science and to see its pertinence to society, perhaps enrollment in physics would increase. We do not seem to have any indication as to which courses over time prove most attractive to student clients. In short, beside the grading practice it may be that the content and the mode of presentation all go to influence the student "mystique" concerning physics. "Hockey (111) in his analysis of the Nuffield Physical Science Project concluded that student involvement in individual projects appeared to be a positive factor in the Nuffield program.

In an attempt to find out what is involved in learning and remembering physics, Shavelson (241) gave aptitude and word association tests, in addition to a pre-test, to a group of 28 male students who received physics instruction. Twelve controls got all the tests but no instruction. He found that abstract reasoning ability predicted final scores after five two-hour physics sessions but not pre-test scores. What would account for this difference? Shavelson surmised that students who performed well on his post test probably got the relevant concepts early in the instruction. If such were the case, those students would have more opportunity for rehearsal and application of the concepts. Succeeding ideas presented by the teacher would be relatively less novel.

It may be that physics concepts, being as heavily loaded with connections to other ideas as they are, present some rather unique learning problems that ought to be studied in more depth. Norwood
Russell Hanson in *Patterns of Discovery* refers to physics terms or concepts as being heavily theory-laden. Thus, what we may be trying to do without knowing how is to teach a complex of relationships connected with each major concept. In contrast to less highly integrated disciplines, these networks of relationships accumulate in successively more complex patterns. We do not know how to help people enter that network. A few people do it naturally, but the rest will continue to suffer until we understand more about the nature of learning in different subject matter areas.

For schools and colleges out to save money and perhaps time, Kemper's (136) comparison of semi-automated and traditional laboratories merits some attention. Students in both groups did equally well on measures of achievement and quality of laboratory reports.

Lerch (159) found that college students who took part in what he called divergent laboratory experiences in physics met more of the objectives outlined by the Commission on College Physics than did those in what was presumably a traditionally convergent laboratory. He failed to find out which kind of laboratory nurtured a taste for physics in what kinds of students.

From his study of college students enrolled in a physics sequence for non-majors, Spoeti (258) found that major themes such as symmetry and conservation appeared to produce desirable achievement. In this approach he concurred with the staff of the Conceptually Oriented Program in Elementary Science (COPES).

In the United States it appears that we have a fairly well trained corps of physics teachers who like to teach in the high direct mode and who give low grades. For students they have college bound types, most of whom are in the course by necessity rather than by choice. According to the Ivany et al. (120) survey, teachers present both classical physics and the development of theoretical models. Physics programs developed with government subsidy are fairly widely disseminated.

In Australia where PSSC has had considerable dissemination, the picture appears to be similar [Gardner (82), Goodwin (89), MacKay (165)]. The tenor of these reports generally suggests that the introduction of this new curriculum has produced some changes in the way teachers conduct instruction. However, the laboratory facets of PSSC are not so well incorporated as the authors would wish. Goodwin (89) reported that there is a core set of topics and then electives which teachers can use in physics. Teachers usually choose electives that build on the core and thus increase the chances that students will pass the state examinations. In doing their planning for physics instruction, colleges can only assume the core set of units.

In the United States, Elizabeth Wood (301) suggested that the National Assessment Examination be used as a pre-test in physics so that teachers will know what they have to teach. The examination

---

Science and the Handicapped

In this section we discuss the relation of education in science to two groups of children with physical deficits, the deaf and the blind, and to a third group of students, those who come from socioeconomic circumstances that appear to have handicapping consequences as far as school performance goes.

Boyd and George (30) found that in a ten-week period deaf children (ages 10-13) could learn to categorize objects according to their properties. The children who took part in activities similar to those from the SCIS and SAPA elementary science programs improved substantially their ability to categorize. This is in contrast to what has been the case for many normal groups which do not especially benefit from the early categorizing activities that characterize many of the modern elementary science programs. Presumably children who are physically normal learn categorizing behaviors through sources other than science. The fact that the 26 children in this study have reached 10-13 years of age without acquiring skills normally in the repertoire by 8-9 years suggests an experiential deficit.

Linn and Peterson (160) used SCIS materials to teach classification skills to visually impaired students as well as to what they called culturally diverse and middle class groups. Controls received no instruction. Both the visually impaired and the culturally diverse performed better after instruction than did the controls but the middle class group did not. Although in this study there is no indication that groups were equated on age or aptitude, we nevertheless learn two things. If children do not know something, they can be taught it; and if they already know the material there is not much point in teaching it to them. SCIS personnel have worked closely with a school district in San Jose, California, to adapt the materials of SCIS for use by visually impaired elementary students.

Nancy T. Long (163) tried SCIS materials adapted for the blind with 14 students in Washington, D.C., who ranged in age from 9 to 19. A matched group of controls drawn from San Francisco schools had whatever science program was offered there. The experimental group performed at a higher level on three tests: Histograms, Systems and Interaction, and Science-Learning. Since these three tests are rather specific to the content of the SCIS program and since the kind of science, if any, received by the controls is not described, this study fails to tell us much more than that visually handicapped students can learn SCIS content. Had the controls been in classes where SCIS was being taught then we might at least know how much the adaptations of equipment for the blind helped them to learn.

From the one study on the deaf and two on the blind we infer that such students can learn some simple science. We also infer that direct manipulative experience given much earlier in the development of handicapped children might help to reduce the magnitude of the deficits in
performance which handicapped students often exhibit. In a separate article published earlier (1972) Linn reported that blind children using SCIS made substantial gains in attaining both process and content objectives.

We suggest that careful clinical-type studies of how blind and/or deaf children learn science could cast some light on the learning processes of physically normal children as well.

Rowe (231) suggested that some ghetto children could be thought of as functionally deaf if they were markedly deficient in the language used for communication in schools. Words without meaning fall on deaf ears (Chapters 5, 12 and 13). The highly manipulative nature of elementary science programs might be the means for bridging the communication gap, if indeed one exists. Cooper (57), for example, found that Latin American students enrolled in physics at the University of Texas achieved higher scores when taught in Spanish than when taught in English. He also found that a good deal of English could be introduced without impairing achievement in physics.

Kral (146) introduced ESS (Elementary Science Study) units as a supplemental science program to his experimental group and provided a placebo consisting of films, current events and discussions to his control group. He found that both Indian and non-Indian students in the experimental group scored higher on the standard achievement test than did their control counterparts. Students from higher socio-economic backgrounds benefited more than did those classified as low socio-economic. Non-Indians scored higher than did Indians. Had Kral used a multiple regression procedure, we might have learned how much variance on the achievement measure was contributed by each of the background variables. Such information would have much more diagnostic use than the set of 2 x 2 factorials he employed to compare groups.

In an interesting pilot venture, Elizabeth Wood (302) rewrote two exercises from the National Assessment pool of items and tried them out with both inner city and affluent groups. She found that the patterns of answers were influenced by the form of the questions. She also suggested that items could be set up in such a way as to yield more diagnostic information. While the size of her sample was too small to make any generalizations, there is at least a suggestion in this study that National Assessment items may be stated in a way that helps one group more than another.

In a neatly designed investigation, Huff and Languis (118) showed that exposure to SAPA, Part A, improved the oral communication skills of disadvantaged kindergartners. The early levels of modern science programs put emphasis on classification. Disadvantaged children apparently benefit from this experience but middle class children usually already can classify and show no distinctive gains.

---

Riechard (224) used a regression procedure to relate variables of inner urban, outer urban and rural farm kindergartners to performance on a Life Science Concept Acquisition Test. I.Q. and chronological age were the best academic predictors while the mother's years of schooling was the best socio-cultural predictor. Time and change concepts were a big source of variation. In this study as in others, we find that knowledge of non-disciplinary concepts such as "opposite, under, before, different, change" are keys to performance.

Johnson (125) did a somewhat similar study in which he related the categorizing performance of high and low socio-economic status kindergartners to background variables. The poor did not perform as well and exhibited more perceptual dominance in their categorizing processes.

Case and Fry (48) found that low socio-economic 14 year olds could learn to design and conduct experiments and to evaluate outcomes if they were given the chance. Using a demonstration interview technique, Donaldson (67) examined 64 Headstart kindergartners and nursery school boys and girls (half of whom were black) regarding their understanding of rockets, seeds, human growth, evaporation and electricity. She found no differences between races but significant differences between members of different socio-economic status. Again we seem to see evidence of a deficit stemming from the environment. However, we cannot be entirely comfortable with this interpretation because she reported that boys did better than girls. This raises a question: is the effect of differential cultural conditioning of the sexes equivalent to forcing a deficit environment on one of the sexes? Donaldson also made some interesting observations on the non-verbal behavior of the children. Rockets provoked more non-verbal expression than did other topics. The children in her sample knew most about electricity and least about evaporation.

Ayers and Ayers (18) studied the impact of SAPA on the rate at which 20 Appalachian kindergartners acquired conservation of numbers, liquid, volume, solids, lengths, weight and area. Since their 20 controls who were not exposed to SAPA had not developed as extensive conservation behavior, they concluded that the rate of attainment was influenced by the program.

David Allen (8) evaluated a program at the Franklin Institute in Philadelphia where students from pairs of schools having different racial and socio-economic backgrounds learned science together one day a week. Thirty schools participated. He examined student interactions, did sociometric studies, gave achievement tests and semantic differentials. In some cases parents participated. There was a general increase in sensitivity to environmental problems but not much evidence that the mixing of students in this kind of science setting for only six days did much for student attitudes.

Acting on the assumption that the out-of-school science interest of children might form a basis for building a school program, Mathanikal (175) sought to discover the preference patterns by sex and by race of pupils in grades seven and eight. On the interest inventory, principal component analysis followed by varimax rotation led to the identification of nine independent factors: academic, nature study, mechanical
hobby, biology experiment, drugs, cosmology, environmental, and high verbal. Whites, Matchanikal noted, preferred activities inspired by inquiry and experimentation. Blacks preferred the more traditional academic types of activities. He noted that expected lower interest of girls simply did not occur. Girls expressed a high degree of participation. A few traditionally sex-associated interests were reversed in this study.

Shipe (247) concluded in his study that non-college bound students in Connecticut did not adequately learn the major ideas in biology. With I.Q. controlled, he reported that students in traditional as compared with BSCS biology were more likely to catch the big ideas. Males were more likely to learn what he called "principles" than were females.

In an examination of individual problem solving behavior, Gilbert (86) found that boys in both low and middle socio-economic groups attempted to explore more alternatives than did girls. More of the disadvantaged children seemed to be frightened and frustrated by the task which consisted of retrieving a piece of candy visible under four stacked concrete blocks. Here we again find ourselves coming to the conclusion that the group of children who most need science in the early years of school are the socio-economic and culturally handicapped.

Rowe (230) reported in a symposium on women and science held at the international science convention in Mexico City in June of 1973, that girls who had taken part in SCIS for several years performed with science materials in the same manner as did boys. She also interviewed 12 female graduating seniors with majors in physical sciences to determine what difficulties they encountered as minority members of their departments. All 12 had jobs in industry and were interviewed separately. All 12 described protracted jibing behavior from teaching assistants as well as from certain professors. Possibly the more difficult slashes came from their sorority classmates who constantly charged them with being in physical science because it was a good place to catch a man. All 12 said that the cumulative impact was to force them to become more self-reliant, to ask fewer questions, and to find their own solutions to problems.

Jackson (121) studied seven predominantly black colleges to find out what institutional characteristics affect learning outcomes of prospective science teachers. She reported that the quality of teaching and faculty-student relationships as well as student perception of the level of scholarship demanded related to performance.

We then come to conclude that children from minority groups seem to benefit from intervention programs; females perform as though they come from a handicapped group. Clearly we need more careful diagnostic assessment to learn how best to help children in these groups. Possibly the part which institutional expectation for performance plays also ought to be examined.
Education, Characteristics and Behaviors of Teachers

The number of studies devoted to research on teachers and their education has more than doubled since last year. The 60 or more studies reviewed fall into clusters. In the teacher education group, ten appraised the effects of special projects such as Academic Year Institutes and School-College Cooperatives (both NSF funded categories of inservice education) on teacher behavior or attitudes.

The second group of investigations focused on personality, roles, and/or behaviors of teachers. Approximately 15 studies examined the relation of teacher behavior to student achievement. Many of these utilized classroom interaction observations to analyze teacher and pupil communication. Another cluster focused on affective aspects of the teaching process.

Teacher Education

While many of the studies reviewed lack some rigor, they do suggest that novel training experiences produce some changes in teacher perceptions of one kind or another. Unfortunately, they do not show us whether these are temporary perturbations or how action in the classroom relates to a particular set of attitudes or perceptions. Neither do they indicate what ingredients in the training might account for specified changes. This kind of analysis, even post-hoc, might eventually help us uncover some persistent patterns.

Six studies which explored the impact of NSF institutes on participants, are reviewed first.

Berger (27) attempted to distinguish between the impact of training and the impact of experience on teachers' role perceptions. He identified four groups: 1) teachers not exposed to SCIS, 2) teachers not exposed but beginning an institute, 3) teachers who completed a two or four week institute, and 4) teachers who completed a four week SCIS institute and had one year experience teaching the SCIS curriculum. Responses on the Predicted Role Measure (PRM) were categorized into a teacher-oriented cluster, a student-teacher cooperation cluster or a student-oriented cluster.

Multivariate analysis revealed no significant differences on PRM scores between the no-SCIS instruction group and the pre-SCIS instruction group. Multiple contrasts intervals indicated that the significant differences found between pre-institute (1 and 2) and past institute groups (3 and 4) were largely due to the teacher-oriented cluster. Teachers with one year's experience teaching SCIS responded on the PMR in the same way as those teachers with SCIS training but with no SCIS teaching experience. In other words, the SCIS-exposed groups differed from the no-SCIS groups, but the contribution of experience to the response scores on the PMR appeared to be minimal.

Bridges et al. (34) evaluated the Cooperative College-School Science Improvement Program to prepare teachers to teach the DISCUS
program. (The DISCUS program is a junior science curriculum designed for and taught to underachievers in Duval County, Florida.) The treatment group of students demonstrated a better understanding of science and showed improved science achievement. The purpose of this study was to show that these effects were not due to treatment alone but were a result of treatment plus teacher training in the CCSS program. Trained teachers compared to non-trained teachers provided more student involvement, less external discipline control and more student-centered instruction. Additionally, trained teachers lectured less and used significantly more concrete image provoking behavior than did non-trained teachers.

Two other studies evaluated the effectiveness of NSF-funded Cooperative College-School Science Programs. In a pretest - posttest design, Pempek and Blick (207) found the project beneficial in changing elementary science teachers' attitudes, particularly those with weak science backgrounds. The Pempek Teacher Behavior Check List assessed change in classroom behavior as viewed by the student while the Pempek Teacher Attitude Scale measured attitudes toward science, science teaching and the scientist. Pempek and Blick sought to connect teacher attitudes with teacher behavior, a laudable objective. Although a significant change in attitudes apparently took place, no significant change in teaching did. They speak of a positive but non-significant change in teacher classroom behavior. A statistically non-significant change is non-significant, regardless of its direction. In other words, the change which they report as positive simply has no meaning in this context.

In contrast to the Pempek and Blick study where attitude changes were not accompanied by classroom behavior changes, Smith and Smith (255) reported that teacher self-assessment was correlated with student perceptions in a pretest - posttest, control design. Project teachers experienced significant change in their perception of science instruction and their students perceived a change which was greater than that perceived by the control group students. Thus, we have some evidence of transmission of training into the classroom context.

Two descriptive dissertations by Macon (166), in North Carolina, and Dyche (69), in Montana, examined the impact of AYI programs on participant characteristics. Macon simply characterized the teachers before and after institute training. From information on registration forms and questionnaires, he concluded that the AYI experience influenced participants to increase mobility, leadership roles and professional activities and status. Conclusions of this nature are somewhat unwarranted as there are no means of determining whether participants had already entered into a period of transition prior to the institute or whether their aspirations changed as a result of the experience.

Dyche (69), in a better designed study, used a control group in a follow-up study of participants. Data from a questionnaire disclosed that institute teachers versus their non-institute counterparts spent more time in laboratory investigation, changed to an inquiry philosophy and made greater use of BSCS materials. The institute participants also read more scientific journals and gained new insights into teaching as a result of associating with other AYI participants.
Other project-associated research resulted in an array of outcomes. Orgren (204) compared teacher strategies employing the New York State Regents' traditional syllabus and the new 1970 revised Earth Science Syllabus. Using a two-way (group x time) factorial design, Orgren found that teachers who opted to continue with the old syllabus differed significantly from those who volunteered to switch to the new syllabus and from those who had prior experience with the new syllabus in respect to teacher strategies, teacher educational opinion, student achievement and student process performance. When all switched to the new syllabus, there were no significant group differences in student achievement gains. It is interesting to speculate as to why these different entering attitudes of teachers nevertheless failed to produce differences in the performance of students. Moreover the students' ability to employ the processes of science did not improve through exposure to the new syllabus.

Two separate studies by Milson (185) and Almase (11) also revealed that teacher behavior and attitude could be modified by inservice courses. Milson looked for attitude changes in secondary science teachers as a result of a short course on environmental quality. Statistical analysis of questionnaires indicated a significant change in teacher responses to items relating to environmental attitudes and environmental management. Unfortunately he failed to find out whether the teacher attitude changes he obtained resulted in any difference in classroom content introduced or in methods of instruction used by the participants.

Almase (11) assessed the effects of a six weeks inservice biology methods course in Cebu, Philipines. In this inservice project the objective was to modify teacher behavior in the direction of more student involvement. Here we meet an interesting problem in prediction. Which of two performances on the Teaching Situation Reaction Test better predicts what a teacher will do in the classroom? This test proved to be a poor predictor of teacher shift toward indirectness. The Processes of Science Test was a better predictor for change toward a process orientation.

In a pretest - posttest matched-pairs design meant to influence career decisions, Conradson (56) explored the effect of actual classroom experience on attitudes of prospective teachers. The experimental group spent three hours per week for twelve weeks with a master teacher. Thirty-two matched pairs finished the experiment. Classroom experience apparently made an impact. Members of the experimental group gained a realistic view of teaching. She also found career decisions were independent of age, sex, class, or grade point average.

A few researchers investigated the effect of different instructional strategies on teacher attitudes toward science and science teaching. One such study by Mitchell (186) compared three groups. Two of these were instructed in the didactic approach and one, the experimental group, received instruction designed to affect attitudes and perceptions toward science teaching. The author reported that trends in the data seemed to indicate that the experimental group responded positively to instruments measuring the inquiry approach, teacher-pupil rapport and interest in science. However, the differences were not significant, so the direction of the trends can have no meaning at this point.
In a similar study Kennedy (137), using a one group pretest-posttest design, focused on differences in attitudes of preservice teachers who received process approach instruction. Differences in pretest-posttest scores indicated process approach instruction effected changes in attitude toward science instruction in a positive way. Again, it would be nice to know how these changes in attitudes correspond to teacher behavior.

Sabulao (234) was also interested in the effect of process approach instruction for developing attitudes and competencies in Cebu, Philippines. Upon completion of the physical science course, Sabulao found significant improvement in competencies of prospective teachers, but no significant improvement in their attitude toward teacher-pupil relations.

Pizzini (214) analyzed the effect of the Iowa-Upstep I & II programs and found them to be effective in contributing to positive growth in self-concept, science teaching philosophy and attitudes toward educational concepts.

In a correlational study, Kuhn (150) found that students with higher grade point averages generally excelled in science teaching competencies. Positive correlations also existed between set induction competencies and preservice science teachers' attitudes toward science.

Of the remaining studies in teacher education, several propose programs for teacher training. Reports of this nature do not technically qualify as research. Typically the procedures for deciding what should be included in such programs come either from curriculum guidelines or from surveys of experienced teachers or occasionally to be decided by fiat.

Through a survey of 52 classroom biology teachers, Beisenherz and Probst (26) identified 154 techniques and skills to be incorporated into a preservice and inservice program for biology teachers. Subsequently, they organized a technique and skills course that could be pursued in an individual format.

In two studies Qutub (216) and Wall and Qutub (296) surveyed academic preparation of earth science teachers in Massachusetts and Wisconsin. As is almost always the case in such surveys, they declared that preparation was inadequate and recommended changes.

Youngpeter (306) examined the role of science museums in elementary teacher education. As a result of the investigation he prepared a checklist as a guide for people who want to set up a museum or to evaluate an existing one.

Five other reports fall in this general category: Phung (213) prepared a program for training secondary school biology teachers in the Mekong Delta Region of Vietnam; Thollairthil (273) analyzed and evaluated science teaching in Nigerian schools; Dixie J. Jones (126) evaluated secondary education programs at the University of New Mexico; Gary Bates (25) developed a sequence of films, "The Classroom Vignettes," for teacher training; and C. T. Abraham (2) studied the NSTA Book Review Program.
Teacher Verbal Characteristics

The second category of research concerning the teacher deals specifically with how teacher characteristics influence student outcomes. A large part of these investigated teacher behavioral strategies and styles.

Several studies examined teacher verbal behavior and its impact on student responses and achievement. Exemplary of this type are dissertations by Aagaard (1), Novak (200), Anderson (13), and Chassas (49).

Aagaard's (1) primary interest lay in the effect of oral questioning of 14 high school chemistry teachers on student achievement. Four teachers taught in the expository mode and ten in an inquiry mode. Of the latter, five used cognitive memory questions and six used high level questions. Analysis of variance and multiple regression techniques revealed that, with I.Q. partialed out, high-level questioning techniques resulted in significantly higher achievement.

Babikian (19) related teacher question patterns in 8th grade science to student performance. Two methods, one in which the teacher asked and answered questions with students taking notes, and one where teachers asked questions, gave partial answers and students filled in the rest, were compared. Children played a more active role in the second method and also demonstrated higher achievement.

In a study of teacher questioning behavior by Arnold et al. (14), lapse time or pause behavior between teacher-student exchanges averaged two seconds rather than one as reported by Rowe (231). They found, as she did, that whether the questions were complex or simple, the mean pausing time seemed to be the same. Additionally, they found that questions of a specific level elicited responses of a similar nature.

In work which is consistent with the results of wait-time studies reported by Rowe, John Novak (200) focused on the effect of timed student responses on teacher behavior as examined by audiotape. Statistical results indicate that timed responses changed teacher behavior significantly in the direction of less teacher talk and lecture. Novak reported no problem with students giving accurate timed responses which in turn resulted in more overt student involvement in the classroom.

In another study which involved silence as a variable, Anderson (13) found that this factor related significantly to all his outcome variables: namely, scores on The Minnesota Test of Creativity, Metropolitan Achievement Test and Test Your Imagination measures. He examined the relationship between measures of classroom verbal interaction and measures of divergent thought and problem solving. Based on scores on the Pupil Attitude Inventory, ten high, ten middle, and ten low classes were selected for study. Observations were collected using Flander's interaction system. From this, four instructional variables were defined: pupil-initiated talk Mesh, PIM; pupil-directed talk Mesh, PDM; IDEAS defined as the frequency of teacher talk encouraging and expanding students ideas; and the ratio of silence to silence-plus-condition (10R). PIM and IDEAS contributed to scores on the creative performance measure. Anderson concluded that the use of
pupil initiated ideas, teacher talk (IDEAS) and the use of silence and seat work facilitated both academic and creative performance.

Another interesting dissertation employing systematic observation of classroom-teacher interaction was done by Yoeh (305) who explored the effect of class-teacher behavior associated with televised science instruction. He ordered classroom behavior into six levels of cognitive content and five levels of affective content. There was a strong indication that students' performance on a criterion test was related to teacher skills in use of questions and reinforcement of cognitive responses. In this study, which shows some effect of attitudes on behavior, Yoeh found that teachers who strongly integrated T.V. science with classroom science had classes which performed significantly better on achievement tests than had teachers who did not favor T.V. instruction.

Sister Chasas (49) also examined the relationship between teacher-student verbal interaction and critical thinking abilities using Flander's observation system. The study population consisted of 12 Puerto Rican college biology classes. All measuring instruments were translated into Spanish. The types of interaction included lecture, discussion with broad questions, and discussion with narrow questions. The Watson-Glaser Critical Thinking Appraisal and the Nelson Biology Test were outcome measures. She found no significant relationship between critical thinking ability and teaching method. Adjusted mean scores showed discussion with broad based questions to be slightly higher than the other methods. Lecture produced greater achievement than discussion with narrow questions. It would be interesting here to have the number and types of questions quantified.

At other points throughout this review as well as in earlier reviews, the Watson-Glaser Test of Critical Thinking has appeared over and over. Yet it rarely produces useful information for the short term situation in which it is usually used. If the test measures a persistent quality, i.e., has trait characteristics, it is unwarranted in the face of the multitude of evidence to expect it to be responsive to short-term treatments. We see, for example, in the Chasas study, as we might expect, the Nelson Biology Test is sensitive to the treatment. In contrast to the Watson-Glaser Test it has the quality of a state variable.

In high school chemistry and general science classes, Wolfson (300) found that standard achievement scores were related to the teacher's indirect-direct ratio. Students of teachers rating high on indirectness attained higher achievement scores.

In an interesting experiment, Jacobs (122) utilized Flanders interaction as an observational learning experience under the pretense of evaluating a program. Thirteen teachers were trained to record the behaviors of master teachers using a Flander's grid and the Indicators of Quality (I.O.Q.) measure. The master teachers were meant to serve as models for the observers. Thirteen other teachers did not get this experience. The verbal behavior of both groups was then examined. Analysis of pre and post treatment audiotapes showed significant changes in the verbal behavior of the experimental group. This is an exciting study in all but one respect. Noticeably lacking is a reference to the
theoretical framework upon which modeling is based. Although it is essentially a study of observational learning or modeling as described by Bandura\(^{14}\) and others, there is no reference of the theoretical underpinnings of the study. Lack of a conceptual model is the most prominent criticism of the studies reviewed in the teacher section.

In an interesting series of studies of which the one reviewed here is only an aspect, Campbell (46) has developed a procedure for examining the influence of chains of teacher behaviors. Campbell has developed a technology for describing sequences of chains of teacher behaviors. Events in a classroom are not discrete. Certain sequences of moves may be more productive than others. Campbell has provided a means for a more careful analysis of the teaching process. He, also, finds wait time as an influential variable. This technique of macro-analysis allowed him to produce further verification of Roe's work. Along with Chalker\(^{15}\) he found that a population of junior high school teachers gave less response time to poor students. Chalker found that teachers who rate higher on a dogmatism scale exhibit shorter pauses.

Smith (254), who had developed a classroom observation instrument referenced specifically to the Earth Science Curriculum Project, studied three teachers each for two weeks to determine the extent of project usage. This technique is potentially useful as a measure of degree of project implementation.

In evaluating a program it would be desirable to have some measure of its degree of implementation before jumping to hasty conclusions based on achievement or other measures. Smith's instrument is one example of a device meant to be specific for a particular program. We would encourage the addition of similar measures to evaluate studies.

Van Alst (286) investigated the effect of influencing teacher and student expectations on student achievement in high school biology. Treatment consisted of one week's counseling of individual students by guidance counselors and of teachers by researchers. No significant differences were found in any variables, sex, ability level, etc., except possibly for the interaction of study ability \(\times\) teacher ability \(\times\) ability level. Van Alst concluded that influencing expectations was complex and difficult. However, one would not expect a one-shot counseling session to be of very great influence.

Vannan (289) described responses of students in a graduate methods course having to read and critique doctoral dissertations. Response was favorable but students generally lacked background in experimental design and statistical background. There is no indication made as to whether this will influence practice or not.


In an exploratory kind of study, Lawlor and Lawlor (153) sought to examine the cues teachers use to categorize students into high, medium, and low ability groups. They exposed 72 undergraduates to two ten-minute video-tapes of children working science tasks. The preservice teachers were then asked to rank the students on perceived ability.

What is interesting in this study is the degree of agreement among the teachers on which children were high ability and which were low ability. The Lawlors sought to discover what observations led to the ranking and found that the 72 people mostly stated inferences with no connection to specific observations. In short their study raises the question of how judgements are formed and suggests that it may be useful to teach some kind of observation system to people.

Teacher Attitudes and Values

A number of studies in this last group concern the attitudes and values of science teachers. Unfortunately not enough of these studies seek to show whether any correlations exist between the attitudes teachers hold and the way they teach.

In Australia, Tisher and Power (278) designed a study to compare values and practices of experienced teachers and prospective teachers with values specified by the Australian Science Education Project. Findings in a preliminary report showed teachers divided into three groups: those who were congruent with ASEP and valued the problem centered approach, open ended discussions and warm student relations; those who held dissonant ASEP values and wanted a narrowly defined curriculum and a teacher dominated classroom; and those who valued warm student relations and a broad curriculum but felt the curriculum should be teacher dominated.

In a second Australian study, Symington and Hawkins (269) compared the values of Australian and British science teachers using the Nuffield Curriculum. They found both groups valued the ability to observe, to classify data and to communicate ideas.

Instead of investigating values that teachers hold, Bybee (45) compared three groups of students to see which teacher characteristics they valued most. The three groups (advantaged, average, and disadvantaged) did not differ in the things they valued. The most highly valued characteristics were good teacher-student relations and enthusiasm in working with students. Amount of teacher knowledge, organization and planning were relatively unimportant characteristics. Valued least by all groups was the instructional method employed by the teacher.

Shrigley (248) examined the connection between attitudes and knowledge. Correlation between these was found to be low among preservice science teachers. He concluded that variables other than the level of cognitive content were influencing attitudes toward science.

Three investigators attempted to isolate specific traits of science teachers that characterized effective teaching. Main (169) compared science teachers and non-science teachers in relation to teacher
effectiveness. Science teachers were shown to be more reserved, calm, mature, taciturn, self-sufficient and resourceful than were other teachers. Female science teachers demonstrated greater warm-heartedness and outgoingness than did their male counterparts. Unfortunately this study falls short of being useful because the author did not relate personality variables to differential teaching practices or to selective interactions with students of like or different personality configurations.

Taylor (271), in a related study, went to the next step and correlated critical thinking ability with two teaching methods; namely, activity-centered teaching and textbook-centered teaching. Discriminant analysis techniques classified teachers into the two groups. Activity-centered teachers were more self-sufficient, affected by feelings and suspicious than were textbook-centered teachers. Again, as in the Main study, females were found to be more tender minded and conscientious while males were more assertive and experimental. A question raised here is whether there is a preference by males or females for one teaching method or the other. In the cognitive domain, activity-centered teachers scored significantly higher in the ability to recognize inferences and assumptions and, overall, scored higher on the Watson-Glaser Critical Thinking Appraisal. The number of units in science courses correlated highly with scores on the critical thinking measure.

Okey (202) studied the effects of Bloom's mastery teaching strategy on teacher attitude and effectiveness. Teachers in a graduate science methods course were taught mastery teaching strategies which they subsequently implemented into their classrooms. Okey found that the experimental group of teachers demonstrated significantly greater positive attitudes. Pupil posttest results indicated that students of the mastery teaching group scored higher in achievement than did the pupils of the control group.

Lazarowitz (155) developed a technique for determining attitudes of secondary science teachers toward inquiry. He developed and validated two instruments, Inquiry Science Teaching Strategy (ISTS) and the Personal Data Form (PDF), for the study. He found teachers with favorable attitudes toward inquiry tended to be older, female, to hold advanced degrees and to teach specialized classes and to use new curriculum material in the classroom. Interestingly, he found that formal research experience related highly to positive attitudes toward inquiry.

Gessner (85) examined medical school faculty effectiveness by comparing student ratings, class exams and national exams. He found that student ratings and national exams correlated. In other words, teachers whose students performed well on the national exams were also rated high by students. Scores on class exams did not correlate with student ratings on faculty members or with national exam scores.

Jackson (121) raised the question, "What effects do black colleges have on prospective teachers?" by examining the institutional characteristics of four predominantly black colleges. Performance of prospective science teachers related significantly to scholarship, awareness and quality of teaching and faculty-student relationships.
A study by Manning (173) also focused on science faculty members in black colleges. She investigated motivation and job satisfaction and found, among other things, that the colleges recruited predominantly from their own geographic area. Most teachers had some other work experience prior to taking a faculty position. Blacks comprised less than 50 percent of the science faculty in 67 predominantly black institutions. The primary interest of most of these faculty members lay in classroom teaching rather than in research. Most were satisfied with their positions and expected to continue.

Studies by Al-Ani (7), Bowman (29), Brett (32), DeGroote (63), Froesch (81), Geisert (84), Hoagland (110), Traughber (282) and Walker (292) were also reviewed.

Values and Philosophy

In a paper, "Social Origins of American Scientists and Scholars," Kenneth Hardy16 provided data to show what cultural influences (regional, religious, school type, etc.) operate to produce larger numbers of productive scientists. He presented a table of contrasting values which distinguish high production from low production groups. "These values probably undergird the cognitive and motivational orientations of the individuals..." (p. 503). People of middle level productivity had less strong positions on each of the values associated with high productivity than did their peers in the high productivity category (see Table II).

If we took the framework of values which seems to be empirically connected to creative output and asked what values do we explicitly or implicitly communicate in our science, what would we find?

Munby (194) developed and tested a scheme in 14 classrooms for getting at some part of the answer to this question. He chose instrumentalism and realism as two views of science and intellectual dependence and independence as values influenced by modes of teaching. The combination of epistemology with the analysis of teaching in which transmission of values are the focus should receive more attention. We found only two other studies that related philosophy of science to the psychological dimensions of values and valuing.

Hudgins (117), who was interested in maximizing man's potential to inquire, made some interesting contrasts between Dewey's pragmatism, a philosophical position, and Freud's conceptions of psycholanalysis. Hudgins maintained that psycholanalysis promotes the kind of self-knowledge that makes a person potentially a more productive inquirer. How to translate this analysis into a series of testable propositions is probably the next problem Hudgins faces.

Obenauf (201) related personalized knowledge and the scientific process in an holistic conceptualization of inquiry. She attempted an answer to the questions, "How do reason and methodology partially

### TABLE II

Cultural Values Associated with High and Low Production of Scholars and Scientists (p. 504, Hardy)

<table>
<thead>
<tr>
<th>High Productivity</th>
<th>Low Productivity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Naturalism.</strong> Belief in a world of order, law, pattern, meaning.</td>
<td><strong>World is unknowable, incomprehensible.</strong> Events are capricious, mysterious, whimsical.</td>
</tr>
<tr>
<td><strong>Intrinsic valuation of learning, knowledge.</strong> To be learned, wise, is highly valued. Broad conception of valued learning.</td>
<td><strong>Suspicion of learning, education.</strong> Constricted view of valued learning. Anti-intellectual.</td>
</tr>
<tr>
<td><strong>Dignity of Man.</strong> Optimism concerning man's ability to discover truth, accomplish things, change the world.</td>
<td><strong>Disparagement of man.</strong> Man is powerless, at the mercy of fate, destiny, luck, chance. He is evil, incompetent.</td>
</tr>
<tr>
<td><strong>Personal dedication.</strong> Seriousness of purpose, sense of mission, positive mysticism. Long range striving. Responsibility beyond family.</td>
<td><strong>Sense of indirection.</strong> Must take, enjoy what is available now. Loyalty to family, kin.</td>
</tr>
<tr>
<td><strong>Antitradiational.</strong> Not satisfied with established ways of doing things. Restless, inquiring spirit.</td>
<td><strong>Traditional.</strong> Past is respected, romanticized. Filial piety valued.</td>
</tr>
<tr>
<td><strong>Centered on near future.</strong> Concerned with this world. Orientation toward the foreseeable future.</td>
<td><strong>Centered on present and distant future.</strong> Hope for a better break in the distant future, the next life.</td>
</tr>
</tbody>
</table>
explain processes by which knowing and the act of discovery occur?" Bridgman's \(^{17}\) personalization of knowledge in physics and Bertalanffy's \(^{18}\) criticism of psychology both occurred to us as we read this dissertation. The groundwork which Obenauf has done could provide a fruitful junction for psychological and philosophical research.

**Surveys**

Since surveys are usually meant to answer questions which are primarily of local interest, we give them little attention in this review. That does not mean they are of little value. On the contrary, they may have considerable importance for planning or ascertaining the state of things so that future plans can be made. Surveys are popular among commonwealth researchers. A few are worth noting because they have wider implications or a quasi-experimental feature to them.

Bonora (28) discussed some of the problems encountered by the International Association for the Evaluation of Achievement as it gathered data in 20 countries. At the international level preliminary analysis focused on identifying family, socioeconomic status, and pedagogical factors which relate to science achievement. Within countries, curricula, attitudes and geographic variables were the chief background variables examined.

R. Hansen and J. Neujar (103) examined the career development of a small sample of males and females in science and found both were equally likely to earn B.A.'s in science but males were more likely to continue work in science or science teaching. In a study of what he called moot beliefs (misconceptions and superstitions) among Wisconsin twelfth-graders, Otto (205) found a significantly greater number of girls who accepted 13 of 36 items, i.e., held misconcepts respecting an item.

A. C. Perkes (209) did a survey of 10th and 12th graders' knowledge of environmental science and attitudes in eleven states of the Midwest and Far West. Twelfth grade students scored better than tenth graders on concepts but not on knowledge of specific facts. There were sex differences in attitudes. Major environmental concerns differed from state to state. Californians worried about air pollution; Wisconsin students, about water pollution; and Hawaiians, about land use.

John Tren't's (283) study of environmental trends as perceived by colleges of education and state departments of education indicated that environmental education is probably on the increase. Fadelli (75) evaluated the conservation content of elementary and secondary school textbooks used in 13 western states and concluded the texts were woefully deficient in conservation knowledge.


\(^{18}\)Bertalanffy, op. cit.
Surveys of science teaching were done by Buckeridge (41) in New England, Mideastern and Southwestern states while Maben (164) did a survey of Central and far Western states. The Buckeridge survey examined certain questions concerning the relation of teacher attendance at NSF institutes to teaching practices. She found that NSF programs influenced teachers' use of various learning activities. In a survey of science teaching in secondary schools of the Great Plains, Rocky Mountain and Southeast region of the United States, for the period 1971-72, Baker (22) found the following: 1) larger schools were more likely to be offering BSCS, PSSC, CHEM study and other programs developed with NSF funds, 2) over 49 percent of the schools used homogeneous grouping, 3) the average annual budget was $1.00 to $2.00 per pupil each for equipment and supplies, 4) instructional techniques most commonly employed were lecture-demonstration, group laboratory instruction, individual laboratory instruction, and 5) NSF summer institute attendance seemed to be a factor in the disposition of teachers to use pupil-centered instructional activities.

Sikes (249) tried to find out how much environmental science was being taught to gifted science students in Texas public schools. He also sought to determine the nature and effectiveness of environmental science teaching. Generally he found that environmental sciences receive little attention in most science classes.

Riggs (225) did a survey of student and instructor attitudes toward biology and Biolabs in two year colleges. There was not much of a relationship between instructor and student attitudes. Attitudes were generally more favorable in courses that combined self-paced and group-paced instruction.

Yao (304) provided a description of science and technical instruction in China during 1949 to 1957 and discussed the relevance of Soviet educational practices to Chinese Communist education. Yao concluded that the uncritical absorption of the Soviet system did not serve the Chinese well, particularly because there was so much emphasis on rote learning.

Rosier (227, 228) reported in two separate articles results of the international survey as it pertained to Australia. Australia ranked third out of 19 countries. He analyzed the particular contributions made by science II (14-year olds) and IV (terminal secondary students). Countries differed in their ability to retain group II in school. Different states in Australia also differed in their holding capability and science performance. He also noted that Australian teachers were generally younger than those in other countries.

Kelley and Nicodemus (133) described early stages in the diffusion of the Nuffield A-level biology. Teachers learned about the program mostly through professional meetings. The decision to adopt was made at the departmental level 46 percent of the time and by individual teachers 52 percent of the time. A one-week workshop to introduce the course increased the chances teachers would use the materials, but teacher characteristics also determined whether the Nuffield biology program would be used.
Bradley and Hutchings (31) used a questionnaire to see how concepts of science and scientists related to subject choices in school in England and Wales. Males were more favorable to science than were females and liked physics and chemistry better than biology. Students saw scientists as family men who were kind, cheerful, interesting and not easily discouraged. Females thought science was too fact-oriented and viewed people in the liberal arts more favorably.

In Israel, Jungwirth (129) found that college students and scientists responded differently on the TOUS. Scientists expressed 85 percent agreement with the images portrayed by the authors of the TOUS. Exposure to BSCS has not succeeded in changing high school students' views to conform with those of scientists.

Soh (259) studied potential scientists' and non-scientists' motivational conditions by administering the School Motivation Analysis Test to secondary students grouped in these two categories according to an occupational scale. Soh concluded that differences can be explained in terms of dynamic traits and that science bias was instrumental in securing ergic gratification.

Jungwirth (128) conducted a study in Israel to find out how the preferences for indoor and outdoor activities of BSCS and non-BSCS students differed. The patterns of preferences for the two groups did not differ. Some sex differences emerged. Females preferred indoor study of plants while males would rather study animals.

Dieter (64) asked each of 220 judges in the 1970 Outstanding Biology Teacher Award program to rate items about teachers' performance. His intention was to establish teacher evaluation criteria. The survey showed distinct biases due to judges' occupations. To counter these biases, Dieter suggested there should be input from teachers in the construction of evaluation procedures and criteria.

A survey of responses by Northern Ireland pupils to the Schools Council Integrated Science Project (SCISP) showed that 55 percent would prefer more biology and that males showed less interest in the social aspects of science than did females.

Other surveys reviewed for this section were conducted by Al-Taiey (12), Gilbert and Mortenson (87), Gough (90), Hackett and Holt (96), Jensen (123), Jones and Roswell (127), Kohn (143), Kraus (147), Leake (156), McKenna (180), Milby (182), Shaw, et al. (243), Summier and Broadhurst (266), and Tomikel (280).
Summary and Conclusions

Now that this review is done, what can we say about the research in general? Of course, one best answers such a question when a particular question arises. For example, is it useful to teach science to elementary children? Is the effort to individualize instruction worth it? What can we do to improve student interest in physics, etc.? A particular question brings to mind relevant findings. Without those questions, a summary may lack some focus. However, we shall try to make some general statements which seem supported by the research. These will necessarily have a kind of advocacy character to them if readers with different interests would ask different questions.

Start with elementary school. The quantity of research in elementary science has fallen off. It appears that most of the science programs, developed in the past decade help language, logic and interests to develop among the disadvantaged and handicapped. These groups, in particular, benefit during the first three years of school by exposure to science that has an activity dimension to it. Middle and upper class primary children show relatively less benefit as the programs now stand. Probably the heavy emphasis on sorting, categorizing, ordering, etc. that marks the early programs makes up for deficits among the poor and handicapped but adds little to what the environment already supplies to the more well-to-do. Thus, we would suggest that children who are most frequently denied science on the grounds they need "fundamentals" should be given more of it, namely, the disadvantaged. Programs suitable for more advanced primary students still need to be developed.

We know relatively little about the cumulative impact of four or five years of elementary science. While many districts have now had four to six or more years experience with coherent describable science programs, almost no research or evaluation is being published that describes the cumulative impact of successive years of science instruction. We need this kind of research to answer questions of worth and appropriateness of science instruction. We also need to know whether the differences in performance and attitudes between the two sexes is minimized by equal exposure over time to a sound science program. The research should also tell us whether the curriculum developers, who complained in earlier years that what they were trying to develop took time, are right. Namely, are there long-term gains of some kind?

A great deal of effort has gone into developing and evaluating self-paced, auto-tutorial programs of various descriptions. There is an apparent paradox here. Most of the researchers at the high school level report favorable results while the outcomes at the college level are much more varied. In general, if you are going to do self-paced instruction, the units should be short. If students are of ordinary ability, the testing should be frequent. At both levels student procrastination is a major problem. More consideration should be given to making the instructional materials available in different formats so that learners with different aptitude patterns find presentations suitable for them. In self-paced and auto-tutorial areas especially, researchers and curriculum developers need to explore the implications of aptitude-treatment-interaction research.
Physics teachers are evidently a species apart. They tend to prefer the didactic mode of instruction. They apparently grade unrealistically in comparison to the rest of the teacher population. The content of physics instruction frequently calls for levels of logical operation which students have for the most part not yet attained. One procedure might be helpful for members of this group. If physics teachers would perform a content and task analysis of the concepts they are trying to teach, they would uncover the structure of the content. Physics consists of a pyramiding of concepts. Some concepts subsume others. It also entails relationships between concepts. Once these two aspects are drawn out in a content or network analysis, one can also determine the types and levels of logical operations which students need to perform to pass successfully through a unit of content. A whole series of specific analyses of the formal structural attributes of physics accompanied by measures of acquisition would collectively allow us to discover the psychological aspects of physics.

What we need, in short, is a fine grain analysis of structure coupled with a careful psychological appraisal of what is involved in acquisition of the structure. George Polya's insightful book, How to Solve It, suggests one instructional procedure that merits investigation in the context of physics: "If you can't solve this problem, solve a simpler one like it." Use of analogies and different instructional formats merits examination. The production of written materials in physics also needs examination. Readability criteria ought to receive more consideration. Complex ideas probably should be presented in short segments marked by different forms of redundancy. Strangely enough, there is little research on learning in physics. We know very little about what is involved in teaching people a hierarchically structured content that draws heavily on formal operational logic.

Teacher education continues to be a major source of study. Most instructional programs include some interaction analyses. The inclusion of this training has apparently increased the flexibility of teachers, i.e., brought more moves under deliberate control. Indirect modes, e.g., in some modification of a Flanders Matrix, have been related positively to development of student inquiry and better problem solving. Pausing rates of teachers appear to influence amount and quality of student verbal output. Direct modes help certain kinds of content acquisition such as basic concepts and mathematical procedures. Different teaching techniques produce some specific student outcomes. For example, warm supportive rewarding behavior appears to help concept learning but also to reduce risk-taking in inquiry. So objectives need to be clear, and the instructional techniques adjusted accordingly. At the moment most of the studies in this field are of such short duration that the consequences of a prolonged exposure to particular techniques is not well understood. How people learn the complex of relationships that is characteristic of science and what they learn in science need more explicit attention from both interaction and learning research people. What values are being transmitted and what kinds of modeling behavior by the teacher or organizations of discussion favor particular kinds of value development?

Despite the rather overwhelming indications that objectives do not help the learner and in some cases may actually be harmful, we continue to see studies on objectives that fail to refine the questions being asked. The burden of the evidence to date suggests that the learner imposes his own organization on the material. In our view, if people must study objectives, they ought to be asking what kinds of objectives are helpful to what categories of learners. There exists some evidence, for example, that objectives may be helpful to low and moderate ability students but harmful for students of middle and high ability. Researchers have not determined whether students use the objectives supplied or how objectives function in relation to the learner. If objectives can be categorized in cognitive and affective groups, perhaps the latter ought to receive more attention.

Very little productive research attention is being given to the development of values, ethics and processes which might be unique outgrowths of science. In the last analyses people resort to measuring concepts and to using global measures of attitudes (e.g., the TOUS) or critical thinking (e.g., the Watson-Glaser). In our opinion we need a set of specific studies tied to specific curricula and we need to examine performances of students with different durations of exposures to those curricula. After we have seen outcomes from a set of specific situations, we can begin to ask what is general, if anything. The recent inclusion of personality variables will probably eventually prove fruitful. At the moment we know that people with certain personality configurations seem to be especially attracted to certain disciplines.

The section on teacher characteristics, behaviors and training constitutes a diversified mix of studies.

A number of researchers were interested in the effect of inservice training programs on teachers. Included are programs such as NSF Academic Year Institutes and Cooperative College - School Science Improvement programs as well as short independent programs. Training of this nature does appear to make a difference in teachers' personal characteristics. Additionally, there is some evidence of transmission of training into the classroom. The research indicates that inservice programs can influence teachers to exhibit better attitudes toward teaching science, to become more indirect, and to use teaching strategies congruent with the goals of the newer curricula. There is some evidence that students are more likely to learn content which is an outgrowth of inservice training.

Several studies report the effects of process approach instruction on attitudes. While most of these show that teachers receiving the instruction have more positive attitudes toward science teaching, there is little indication concerning the relation of attitudes to modes of teaching or success of students. On the contrary, there is some suggestion that the students of teachers who started teaching a new program only reluctantly did as well as the eager early joiners. It can also happen that while competencies may improve as a result of training, attitudes may be negative.
Research involving the classroom teacher dealt largely with teacher behavioral strategies and styles. Repeatedly it was shown that teacher questioning patterns influence students' academic performance. Wait time and patterns of teacher-student interaction also alter student performance.

Several attempts were made to isolate science teacher characteristics as compared to non-science teachers. We really do not see this as a fruitful line of research. In general, science teachers were shown to be more self-sufficient-resourceful and reserved than other teachers. Female science teachers as compared to male science teachers are more tender minded and outgoing. We are prompted to ask rather impolitely, "So what?" Unless one can relate teacher personality variables to classroom teaching practices or to student performance we see this kind of research as fruitless. Instead we would ask, are teachers with particular personality configurations more helpful to students with the same configurations?

In summary, it seems that research has made some headway toward answering some nagging and persistent questions regarding the art of teaching science but there is much to do, especially in the realms of evaluation, learning and transmission of values.

All the conditions exist in nature at the moment to do this research. Some people ought to do it.
<table>
<thead>
<tr>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract Reasoning, 37</td>
</tr>
<tr>
<td>Abstract Thought, 12</td>
</tr>
<tr>
<td>Academic Year Institutes, 44</td>
</tr>
<tr>
<td>Activity vs Text Centered Instruction, 51</td>
</tr>
<tr>
<td>Advance Organizer, 15, 16</td>
</tr>
<tr>
<td>Affective Outcomes, 31, 32</td>
</tr>
<tr>
<td>Affective Variables, 34</td>
</tr>
<tr>
<td>Age, Aging, 12, 13</td>
</tr>
<tr>
<td>Analytic Thinking, 28</td>
</tr>
<tr>
<td>Aptitudes, 6, 13, 14, 18, 20, 23, 24, 25, 26</td>
</tr>
<tr>
<td>Astronomy, 3, 26</td>
</tr>
<tr>
<td>Attitudes, 30, 31, 34, 35, 44, 45, 50, 51</td>
</tr>
<tr>
<td>Audio-tutorial, 2, 15, 16, 20</td>
</tr>
<tr>
<td>Australian Education, 55</td>
</tr>
<tr>
<td>Australian Science Education Project, 50</td>
</tr>
<tr>
<td>Ausubel, 8</td>
</tr>
<tr>
<td>Biology, 15, 18, 19, 20, 22, 26, 31, 46, 48, 55</td>
</tr>
<tr>
<td>Biology Methods, 45</td>
</tr>
<tr>
<td>Black Colleges, 42, 51, 52</td>
</tr>
<tr>
<td>Blacks, 3, 35, 42</td>
</tr>
<tr>
<td>Blind, 39</td>
</tr>
<tr>
<td>BSCS, 25, 33, 42, 44, 45, 55, 56</td>
</tr>
<tr>
<td>Careers, 45, 54</td>
</tr>
<tr>
<td>CCSS, 44</td>
</tr>
<tr>
<td>Chemistry, 12, 15, 21, 22, 23, 26, 29, 31, 35, 48</td>
</tr>
<tr>
<td>CHEM Study, 12, 55</td>
</tr>
<tr>
<td>Chinese Education, 55</td>
</tr>
<tr>
<td>Classification, 12, 28, 40</td>
</tr>
<tr>
<td>Cognitive Conflict, 29</td>
</tr>
<tr>
<td>Cognitive Preference, 32, 33</td>
</tr>
<tr>
<td>Cognitive Style, 10</td>
</tr>
<tr>
<td>Computer Assisted Instruction, 23, 24</td>
</tr>
<tr>
<td>COPES, 6, 30</td>
</tr>
<tr>
<td>Counter Intuitive Events, 11</td>
</tr>
<tr>
<td>Critical Thinking, 29</td>
</tr>
<tr>
<td>Deaf, 39</td>
</tr>
<tr>
<td>Dependence Proneness, 10</td>
</tr>
<tr>
<td>DISCUS, 43</td>
</tr>
<tr>
<td>Dogmatism, 31, 49</td>
</tr>
<tr>
<td>Earth Science, 22, 23, 27, 41, 46</td>
</tr>
<tr>
<td>Education, Cross-Cultural, 36</td>
</tr>
<tr>
<td>Engineering, 10, 11, 20, 22</td>
</tr>
<tr>
<td>Environmental Action, 32</td>
</tr>
<tr>
<td>Environmental Education, 33, 41, 45, 54, 55</td>
</tr>
<tr>
<td>Error Rate, 21</td>
</tr>
<tr>
<td>ES, 27</td>
</tr>
<tr>
<td>ESCP, 49</td>
</tr>
<tr>
<td>Evaluation, 8, 11, 27, 28, 29, 31, 34, 54</td>
</tr>
<tr>
<td>Evolution, 4</td>
</tr>
<tr>
<td>Expectations, 49, 50</td>
</tr>
<tr>
<td>Field Theory, 9</td>
</tr>
<tr>
<td>Funding Sources, 4</td>
</tr>
<tr>
<td>Genetics, 7, 8, 13, 22, 32</td>
</tr>
<tr>
<td>Geology, 23</td>
</tr>
<tr>
<td>Grading, 27, 31, 36</td>
</tr>
<tr>
<td>Group Process Variables, 10, 11</td>
</tr>
<tr>
<td>Handicapped, 2, 13, 39</td>
</tr>
<tr>
<td>HEADSTART, 41</td>
</tr>
<tr>
<td>Hierarchies, 7, 12, 21</td>
</tr>
<tr>
<td>Homework, 11</td>
</tr>
<tr>
<td>Hunches, 9</td>
</tr>
<tr>
<td>Hypothesis Formation, 20</td>
</tr>
<tr>
<td>Implementation, 49</td>
</tr>
<tr>
<td>Indians, 40</td>
</tr>
<tr>
<td>Individualized, 18</td>
</tr>
<tr>
<td>Information Density, 7</td>
</tr>
<tr>
<td>Information Processing, 23</td>
</tr>
<tr>
<td>Information Theory, 6</td>
</tr>
<tr>
<td>Inner City, 29</td>
</tr>
<tr>
<td>Inquiry, 34, 51</td>
</tr>
<tr>
<td>IFI, 22</td>
</tr>
<tr>
<td>ISCS, 19, 21, 30</td>
</tr>
<tr>
<td>Listening Versus Reading, 19</td>
</tr>
<tr>
<td>Logical Operations, test of, 13</td>
</tr>
<tr>
<td>Mastery Teaching, 51</td>
</tr>
<tr>
<td>Medical Education, 51</td>
</tr>
<tr>
<td>Museums, 46</td>
</tr>
<tr>
<td>National Assessment, 3, 38</td>
</tr>
<tr>
<td>Nuffield Biology, 55</td>
</tr>
<tr>
<td>Nuffield Science, 29, 37, 50</td>
</tr>
<tr>
<td>Objectives, 14, 15, 16</td>
</tr>
<tr>
<td>Occupational Success, 10</td>
</tr>
<tr>
<td>Participation, 11</td>
</tr>
<tr>
<td>Personality Variables, 19, 20, 22, 31, 35, 50, 51</td>
</tr>
<tr>
<td>Personalized Knowledge, 52</td>
</tr>
</tbody>
</table>
Philosophy, 52
Physical Science, 30
Physics, 2, 14, 15, 21, 26, 32, 34, 36, 40
Pictures, 25
Planetarium, 26
Preferences, 33, 34, 41, 56
Problem-Solving, 42
Process Skills, 35
Procrastination, 23
PSSC, 15, 38, 55
Questioning, 11
Reading, 5, 24, 28
Retention, 28
Risk-taking, 9
Rote Learning, 33
SAPA, 14, 27, 40, 41
Scalogram Analysis, 12
SCIS, 13, 27, 28, 39, 42, 43
Self Assessment, 35
Self-Concept, 33
Seriation, 12
Socio-economic Variables, 40
Statistical Significance, 3
Structure, 7, 9, 12, 21
Student Perception, 50
Teacher Characteristics, 46
Teacher Competencies, 46
Television Science, 48
Test Construction, 34
Testa, 27, 51
Time Sense, 8
Urban and Rural Differences, 41
Values, 34, 50, 52
Wait-time, 47
Women, 42
Zoology, 29
References

Citations containing ED information indicate documents available from ERIC Document Reproduction Service, P.O. Box 190, Arlington, Virginia 22210.


98. Hall, John Franz. "The Use of History of Science Case Studies with First Year Education Students to Teach Skills Involved in Scientific Thinking." University Microfilms, Ann Arbor, Michigan, 1972.


196 pp.


