DOCUMENT RESUME

ED 223 475 SE 040 063

AUTHOR	Anderson, Ronald D.; And Others
TITLE	Science Meta-Analysis Project: Volume I. Final Report.
INSTITUTION	Colorado Univ., Boulder.
SPONS AGENCY	National Science Foundation, Washington, D.C.
PUB DATE	Dec 82
GRANT	NSF-SED-80-12310
NOTE	406p.; For related document, see SE 040 064. Contains occasional light and broken type. Produced by the Laboratory for Research in Science and Mathematics Education.
AVAILABLE FROM	Laboratory for Research in Science and Mathematics Education, c/o Dr. Ronald D. Anderson, Campus Box 249, University of Colorado, Boulder, CO 80309.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC17 Plus Postage.
DESCRIPTORS	*Academic Achievement; Advance Organizers; Computer Assisted Instruction; Elementary School Science; Elementary Secondary Education; Individualized Instruction; Inquiry; Mastery Learning; Programed Instruction; *Science Course Improvement Projects; *Science Curriculum; Science Education; *Science Instruction; Secondary School Science; *Teaching Methods
IDENTIFIERS	*Meta Analysis; National Science Foundation; *Science Education Research

ABSTRACT

The National Science Foundation funded a project to: (1) identify major areas of science education research in which sufficient studies have been conducted to permit useful generalizations for educational practice; (2) conduct meta-analyses of each of these areas; and (3) prepare a compendium of these meta-analyses along with interpretative and integrative statements. This report constitutes volume I of the compendium. Four separate studies are reported following introductory comments on the project. These studies and authors are: (1) "The Effects of New Science Curricula on Student Performance" (James A. Shymansky, William C. Kyle, Jr. and Jennifer M. Alport); (2) "Instructional Systems in Science Education" (John B. Willett and June J. M. Yamashita); (3)"The Effects of Various Science Teaching Strategies on Achievement" (Keven C. Wise and James R. Okey); and (4) "The Effect of Inquiry Teaching and Advance Organizers upon Student Outcomes in Science" (Gerald W. Lott). Each study includes a separate table of contents, purpose, methodology, results, conclusions, and supporting documentation. (Author/JN)

Volume I

SCIENCE META-ANALYSIS PROJECT: FINAL REPORT OF NSF PROJECT NO. SED 80-12310

Ronald D. Anderson

Stuart R. Kahl

Gene V. Glass

Mary Lee Smith

M. Lynnette Fleming

Mark R. Malone

Laboratory for Research in Science and Mathematics Education

University of Colorado

Boulder, Colorado 80309

December, 1982

This material is based upon work supported by the National Science Foundation under Grant No. SED 80-12310. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

TABLE OF CONTENTS

Appendices

PREFACE

Science education probably has the longest and richest tradition of research of all the subject fields in education. With a background in one of the natural sciences, it is not surprising that university personnel in science education would have been interested in empirical studies. The interest was strong enough that over a half century ago they formed their own research organization, long before analogous research organizations were commonplace in other subject areas. The interest persists; well over 3000 dissertation studies alone have been conducted in science education since mid-century. Science educators have been interested in conducting research and in examining its implications for classroom practice.

But when translating research results into practice, science educators have faced the same difficulties encountered by other educational researchers and scholars in all areas of social research, namely, how can you integrate the many findings acquired from varying research settings and having conclusions with less than perfect agreement. The numerous variables and less than perfectly controlled situations common to all social research have left science education with the task of finding meaning in a complex set of research findings.

Quantitative procedures for integrating research findings give hope that the extent body of research literature can be given increased meaning and clearer implications for practice.

The recent emergence of meta-analysis led to its application in several places in science education and finally to a proposal to the National Science Foundation seeking support for the project reported herein. Its purpose was to (1) identify the major areas of science education research in which sufficient studies have been conducted to permit useful generalizations for educational practice, (2) conduct meta-analysis of each of these areas, and (3) prepare a compendium of these meta-analyses along with interpretive and integrative statements. This report constitutes that compendium.

A project of this scope, of course, involves a large number of people and acknowledgment of their extensive efforts is gratefully given. Although varied in terms of their role and involvement, each made important contributions to the overall endeavor.

The local project staff included Stuart R. Kahl, who served as associate director during its first year. His work in coordinating the literature searches and coding form development, as well as in preparing common data file formats and serving as a statistical consultant to the research assistants, was of major importance. Other local staff included Gene V Glass and Mary Lee Smith, developers of the meta-analysis technique, who were a guiding force in the development of the project, trained the research assistants, and served as consultants to them in their work. Other key people in the project were the secretaries, Ellen Ward and later Lisa Hamilton, whose sterling performances were invaluable.

The staff extended far beyond the University of Colorado and included personnel from six other universities. The original participants were an advisory committee consisting of one person from each institution as follows:

J. Myron Atkin, Stanford University
Robert Howe, Ohio State University
James Okey, University of Georgia
Lee Shulman, Michigan State University
James Shymansky, University of Iowa
Wayne Welch, University of Minnesota

All played an important role in shaping the project at its inception; two of them, James Okey and James Shymansky, continued in the project as researchers.

The extensive work of the many researchers involved in this project is reflected in their authorship of the several chapters of this report. In addition to persons already mentioned they include a cadre of people appointed as research assistants at each of the institutions involved. Among this staff were William C. Kyle, Jr. and Jennifer M. Alport of the University of Iowa, John B. Willet and June J. M. Yamashita of Stanford University, Keven C. Wise of the University of Georgia, Gerald W. Lott of Michigan State University, Gary L. Sweitzer of Ohio State University, Cynthia Ann Druva of the University of Minnesota and Mark R. Malone and M. Lynnette Fleming of the University of Colorado. Their work reflects not only the many hours required in the labor-intensive meta-analysis process but their high level of professional competence and scholarly ability.

RDA

INTRODUCTION TO THE SCIENCE EDUCATION META-ANALYSIS PROJECT

Ronald D. Anderson

Stuart R. Kahl

Gene V. Glass

Mary Lee Smith

Laboratory For Research in Science
Mathematics Education

University of Colorado

Boulder, Colorado 80309

SCIENCE EDUCATION: A META-ANALYSIS OF MAJOR QUESTIONS

While meta-analysis (Glass, 1976) has been on the educational research scence for only a few years, it has become established as an important technique. It is proving useful in translating the results of numerous studies on a particular topic into a concise form that is reflective of the multiplicity of data found in the many studies, and understandable to the educational practitioner who may be in a position to apply the results. The characteristics of this methodology and guidelines for employing it are well documented (Glass, McGaw, and Smith, 1981). While this approach already has been utilized for several science education questions, it has additional potential value if applied to the wide sweep of major science education research questions in a systematic manner. Such an approach requires focusing on the major research questions in the field, giving attention to various subquestions subsumed under each major question and examining common themes that cut across the major questions.

A project of this design was conducted under National Science Foundation Grant No. SED 80-12310. Within the conceptual framework described above, a large number of research studies were integrated with the results providing a basis for interpretive and integrative statements about the major questions addressed in the science education research literature.

A MULTI-INSTITUTIONAL ENDEAVOR

Although primarily conducted at the University of Colorado, major portions of the project work were done under a multi-institutional arrangement involving researchers from six other institutions. A leading researcher from each of these institutions constituted an advisory committee to aid in identifying the research questions pursued and assisted in designing an endeavor encompassing the work of one or more researchers from their home institutions in this project. The actual coding and analysis work was conducted by researchers located at the indicated six research centers and the University of Colorado. At each location an individual or a team of up to three researchers conducted this work.

Prior to beginning this coding and analysis work, all of the researchers attended a week-long session for training and coordination of work. During this time each individual or team developed the initial version of the coding forms with a large percentage of the categories and format in common. This process resulted in a data base which can be examined across research questions.

This multi-institutional approach had both advantages and disadvantages. It was possible to involve a large research group which was not already extant at one institution. It had further advantage of stimulating meta-analysis work in a variety of locations where in many cases it was not already underway. One of the disadvantages was the inability to readily shift manpower among questions as their scope became more clearly identified during the actual coding process. As a result there is variation in the thoroughness with which the literature has been sampled for each of the research questions. Though this variation is identified here as a disadvantage, it is not a serious problem as indicated in a later section of this report.

IDENTIFYING THE RESEARCH QUESTIONS

The first step in the project was to identify the major science education questions to pursue. It was accomplished by a combination of (a) empirical analysis of the extant research, and (b) expert judgment as to the importance of particular questions. Major attention was given to the empirical analysis rather than the expert judgment, however, in that the basic approach was to include whatever empirical analysis showed to be the subject of a substantial number of research investigations.

The first step was initiated by collecting and examining a representative sample of science education research studies. Literature was sampled across time and type of publication and included studies from The Journal of Research in Science Teaching, Science Education, Dissertation Abstracts, and the most recent abstracts of presentations for the National Association for Research in Science Teaching annual convention. About 300 such research articles were sampled, and the major (as well as subsidiary) questions addressed recorded.

The questions collected were then classified into some broad, general categories. Five persons classified separate portions of the questions into categories. These categories, developed independently by each of the five persons, had much in common. The entire group of five then examined the questions and organized them into a simple classification system. It resulted in thirteen general areas encompassing all but a small percentage of studies which neither fit within these thirteen categories nor constituted a meaningful grouping themselves.

The researchers then went back to the literature (including the Curtis digests of Research in Science Education of several decades ago) to see if additional research questions fit within the framework that had been empirically derived. This cross-validation indicated the categories were appropriate.

The next step was to develop a full description of each of these thirteen areas. They were identified by a generic question for each area along with sample subquestions. These sample subquestions were examples of a larger set of such subquestions; they were a representative and not exhaustive set. In addition, definitions of terms, descriptions of some variables, and a limited rationale for considering the questions were provided.

A form was then developed on which responses could be obtained from other science education researchers concerning these categories. Twenty people were mailed a full description of the thirteen areas, a response form, and a cover letter requesting that they be prepared to discuss the material by phone. All twenty people responded to a telephone request for their judgments on the relative importance of these questions and the adequacy of the literature for doing a meta-analysis. While these judgments of the relative importance of the questions were of value, the judgments of the relative importance of the questions were largely subordinated by an empirical search of the total science education research.

Literature searches were conducted on a sampling basis to obtain an estimate of the size of the literature and determine if sufficient studies existed for a meta-analysis of each question. Abbreviated computer searches were conducted using data bases such as ERIC, Dissertation Abstracts, and Social Science Research. The citations obtained then were screened to eliminate those items which were not research publications. Subsequent investigation indicated some problems with the manner in which the computer searches had been conducted, so additional searches were done "by hand" as a check. They were done on a sampling basis using selected annual reviews of science education research and Science Education - A Dissertation Bibliography, a listing of all doctoral dissertations pertaining to science education conducted between 1950 and 1977. These procedures provided a rough estimate of the size of the literature pertaining to each of the thirteen questions.

At this point a two-day conference of the advisory committee was convened to confer with the project staff and produce a final classification of research questions for meta-analysis as well as identify important variables to include when integrating the research for each question.

One of the original questions ("What are the goals and priorities of science education?") was eliminated due to an insufficient number of empirical studies, even though it was ranked high in importance. The other twelve questions were recombined into a broader set of
estions as follows:

I. What are the effects of different curricular programs in science?

II. What are the effects of different instructional systems used in science teaching (e.g. programmed instruction, master learning, departmentalized instruction)?

III. What are the effects of different teaching techniques (e.g. questioning behaviors, wait-time, advance organizers, testing practices)?

IV. What are the effects of different pre-service and in-service teacher education programs and techniques?

V. What are the relationships between science teacher characteristics and teacher behaviors or student outcomes?

VI. What are the relationships between student characteristics and student outcomes in science?

While these six questions as stated were pursued initially, some of them were delimited further when subsequent search activities made it clear that they were too broad to complete within the resources of the project.

THE LITERATURE SEARCH PROCESS

Identifying and collecting the research studies to be part of a meta-analysis is a major step in the total endeavor. This aspect of the project will be described in terms of the (a) limitations placed on the studies to be included, (b) search strategies employed, and (c) variations in the literature covered among the major questions within the total project.

Restrictions on Scope of the Questions

Because of the need to keep the meta-analysis to a manageable size and to maintain some degree of commanality among the studies included under a particular question, the following restrictions were placed on the studies to be included.

1. The studies were limited to those conducted in the context of grades K through 12.
2. The studies included were limited to those conducted within the United States.
3. For questions I-IV, only those with a control group were included.
4. The studies were limited to those published in 1950 or later.

The Search Process

In a departure from many past meta-analyses, it was decided that the search process would begin with dissertations because of the thoroughness with which data are typically reported therein, and because such a large percentage of research studies are conducted within that context. This process of searching dissertations was greatly facilitated by the existence of the previously mentioned bibliography which lists all doctoral dissertations pertaining to science education conducted between 1950 and 1977. This document lists approximately 3,200 science doctoral dissertations; the entire document was systematically examined to identify each potential dissertation which, by title and categorization within the bibliography, appeared to be a potential for the meta-analysis. These approximately 1,000 dissertations were obtained on microfilm from the Science and Mathematics ERIC center at Ohio State University. Each dissertation was read to determine if it actually pertained to the topic at hand and, if so, it was utilized in the meta-analysis.

Another facet of the search process was screening the bibliographies in each coded publication to identify additional studies to be included in the meta-analysis. In addition to identifying journal articles through this standard bibliographic search method, ERIC searches and simple screening of the entire collection of issues for the relevant years of selected journals were conducted. Among the various research sites, the procedures for identifying journal reports to be included varied considerably. Whatever mechanism was used, a high percentage of the articles located were reports of studies already coded from dissertations. Finally, some studies utilized in this meta-analysis were reported in other sources such as books or unpublished reports.

Variations in Literature Covered

While there was considerable variation in the amount of literature covered among the several research sites, there was consistency in removing many studies from consideration without coding them once they had been read and their exact character ascertained. While 769 studies were coded, nearly 2,000 studies were read in the process. Among the reasons for excluding studies were the following.

a. The most common reason for eliminating a study was inadequate reporting, i.e., not enough information was provided to make it possible to calculate an effect size.

b. The study did not utilize a control group.

c. The study was not within the K-12 limit; most studies eliminated were college level.

d. The study was conducted outside the United States.

Even given this limiting of the studies included, many of the researchers were faced with a body of literature larger than was possible for them to code and analyze completely within their time limitations. The means of limiting the number of studies varied from one site to another but generally were one of the following three approaches. (a) Some sites found it possible to code and analyze essentially the entire body of literature located through the search procedure described above and contained within the boundaries cited earlier. (b) Some sites chose to limit the scope of their original question to one or more key subquestions. (c) Some maintained the scope of their coverage but selected only a portion of the studies for analysis.

CODING THE STUDIES

Meta-analysis endeavors are very labor-intensive; the most time consuming part is reading each study and recording on the coding sheets each relevant piece of information. Of the dozens of items of information potentially available for a given study, the major one is an effect size that provides a quantitative comparison of the effects of the experimental and control group or in the case of a correlational study, the correlation between two variables). For an experimental study, an effect size is calculated which provides a normalized measure of the difference in performance of the two groups with respect to a specified dependent variable such as achievement, attitude toward science, or any other outcome variable. Symbolized by the Greek letter Δ and abbreviated E.S., effect size is defined as the mean difference between the given variable for the experimental group and control group divided by the standard deviation of the control group.

$$\Delta = \frac{\overline{X}_+ - \overline{X}_c}{S_c}$$

where $\overline{X}_+$ = mean of experimental group,

$\overline{X}_c$ = mean of control group, and

S_c = standard deviation of the control group.

The calculations involved in determining the effect size vary considerably depending upon the particular form of the data reported in a given study. The numerous procedures required in the various situations are well developed (Glass, McGaw, and Smith, 1981).

INTEGRATING THE RESULTS

Once the coding (recording information on all demographic, independent and dependent variables available in the report) for all of the studies in the meta-analysis has been completed, attention is turned to integrating this information. This step involved calculating an <u>average effect size</u> ($\overline{\Delta}$, a simple arithmetic average) from all those obtained on a given outcome variable such as achievement (and/or some particular category of achievement), attitude toward science, laboratory skills or whatever outcome variable has been examined within some subset of the studies involved. Furthermore an average effect size can be calculated for a particular outcome variable from all studies with a particular independent variable and this average effect size then can be compared to the average effect size on the same outcome variable for those studies having a different independent variable. For example, in the meta-analysis of studies of instructional systems in science (at K-12 levels) the average effect size on cognitive achievement for 5 studies of audio-tutorial systems was .09 standard deviations higher than the control groups, while the average effect size on cognitive achievement for 7 studies of "Keller Plan" systems was .49 standard deviations higher than their control groups. This same type of comparison can also be made for other outcome variables. For example, one of the audio-tutorial studies had an affective measure, it was an effect size of .33 in favor of the experimental group. Two "Keller Plan" studies had an affective measure with an average effect size of .52. Similar statements can be made about these two instructional systems with respect to any other outcome measures included in some of the studies and similar comparisons can be made with other instructional systems with respect to any outcome measures included in studies of these systems.

A variety of issues have been raised about the interpretation of such results as described above. For a discussion of the issues the reader is referred to a recent article (Glass, 1982) or book on the topic (Glass, McGaw, and Smith, 1981).

PROJECT RESULTS

The results of the meta-analysis in this project are reported in the following chapters of this report. They include one chapter associated with each of the previously identified questions (two chapters in the case of question III) and a chapter dealing with research issues for which data is drawn from one or more of the separate meta-analyses. Brief descriptions of the data files acquired are provided in each of the individual research papers. Copies of the coding sheets used and the complete bibliography of research studies coded are provided in the appendices of this report of the project. The total data base has been compiled on one master file at the University of Colorado and is available, along with a User's Manual (Kahl, Anderson, 1982), to other researchers who wish to use it.

REFERENCES

Glass, G.V. Primary, Secondary, and Meta-Analysis of Research. Educational Researcher, 1976, 5(10), 3-8.

Glass, G.V., McGaw, B.V. & Smith, M.L. Meta-Analysis in Social Research. Beverly Hills, California: Sage Publications, 1981.

Kahl, S.R., & Anderson, R.D. Science Meta-Analysis Project: User's Guide for the Machine-Readable Raw Data File. Boulder, Colorado: Laboratory for Research in Science and Mathematics Education, University of Colorado, 1982, 138 pages.

University Microfilms International. Science Education - A Dissertation Bibliography. Ann Arbor, Michigan: Dissertation Publishing, University Microfilms International, 1978.

THE EFFECTS OF NEW SCIENCE CURRICULA
ON STUDENT PERFORMANCE

by

James A. Shymansky, William C. Kyle, Jr.
Science Education Center
The University of Iowa
Iowa City, Iowa 52242-1478

and

Jennifer M. Alport
Department of Biological Sciences
University of Natal
Natal, Republic of South Africa 4001

THE EFFECTS OF NEW SCIENCE CURRICULA ON STUDENT PERFORMANCE

An Abstract

Elementary, junior high and secondary school science experienced a tremendous curriculum development and growth beginning in the late 1950's, through the early 1970's, that can be described only as phenomenal. Several groups of concerned scientists and educators developed modern science programs with a major emphasis on the nature, structure, and unity of science while accentuating the investigative, exploratory phases of science, and the development of scientific inquiry. In contrast to these new curricula, "traditional" courses generally tended to concentrate on the knowledge of scientific facts, laws, theories, and technological applications (Haney, 1966; Klopfer, 1971; Schwab, 1963).

The public became very science and technology conscious following the historic launching of Sputnik I by the Soviet Union on October 4, 1957. The numerous "alphabet-soup" curricula which were developed as a result of public outcries and financial support from federal agencies and private foundations were aimed at rekindling student interest in science and upgrading the lethargic science curriculum in the schools. Morris Shamos, a noted physicist, science educator, and curriculum director, estimated that 5 billion dollars were spent to improve K-12 science education during the 15 years following Sputnik I

(Yager, 1981a). A substantial amount of this support was from the National Science Foundation (NSF).

Since the inception of the NSF sponsored curriculum development era there have been numerous evaluation efforts to assess the impact of the new science curricula versus traditional science courses. The question as to whether the newly developed curricula were any "better" than the traditional courses became a leading issue in science education. The large body of research on the effects of the new curricula is generally viewed as inconclusive. A brief scan through the literature reveals that some studies claim that the new curricula facilitate cognitive and/or affective achievement while others claim that they do not. Thus, after 25 years of sporadic implementation, the question of how effective new science curricula actually are in enhancing student performance is still unanswered.

This study utilizes the quantitative synthesis perspective to research integration known as *meta-analysis* (Glass, 1976) to synthesize the results of 105 experimental studies involving 45,626 students. Thus, this study is a quantitative synthesis of the retrievable primary research focusing on the effects of new science curricula on student performance. A total of 27 new science curricula involving one or more measures of student performance are included in this meta-analysis.

Data were collected for 18 *a priori* selected student performance measures. These 18 criterion variables were grouped into 6 criterion clusters as follows:

1. General Achievement Cluster
    - a. Cognitive - low
    - b. Cognitive - high
    - c. Cognitive - mixed/general achievement
2. Perceptions Cluster
    - a. Affective - attitude toward subject
    - b. Affective - attitude toward science
    - c. Affective - attitude toward procedure/methodology
    - d. Self-concept
3. Process Skills Cluster
    - a. Process skills
    - b. Methods of science
4. Analytic Skills Cluster
    - a. Critical thinking
    - b. Problem solving
5. Related Skills Cluster
    - a. Reading
    - b. Mathematics
    - c. Social studies
    - d. Communication skills
6. Miscellaneous
    - a. Creativity
    - b. Logical thinking (Piagetian)
    - c. Spatial relations (Piagetian)

In addressing the overall question of new science curriculum effectiveness, the data are arranged in three broad categories: curriculum characteristics, student or teacher factors, and study design features. The variable analyzed in all cases is student performance measured in terms of the meta-analysis common metric known as *effect size* (Glass, 1976). The effect size is a common metric derived from the various tests of student performance in all the studies analyzed and provides a basis for comparison across the many studies addressing the broad question of curriculum effectiveness.

The results of this meta-analysis reveal definite positive patterns of student performance in new science curricula. Across all new science curricula analyzed, students exposed to new science curricula performed better than their traditional counterparts in achievement, analytic skills, process skills, and related skills, while developing a more positive attitude toward science. On a composite basis, the average student in new science curricula exceeded the performance of 63% of the students in traditional science courses.

Further breakdowns of the student performance data reveal other interesting characteristics of new science curricula. For example, new science curricula in biology (i.e., the BSCS programs) produced the most positive performance scores among the science disciplines, while chemistry and earth science curricula appear to have had the least positive impact. Also, studies involving new science curricula judged to have a low emphasis on laboratory activity showed students out-performing their traditional course counterparts by larger margins overall than those new science curricula judged to have a high laboratory emphasis. On the other hand, studies involving new science curricula judged to have a high emphasis on process skill development showed students out-performing traditional course students by larger margins on analytic skill measures than those involving curricula judged to have a low process skill emphasis.

In terms of overall performance, science curricula produced equally positive results when broken down by grade level (K-6, 7-9, 10-12, post secondary). However, student performance in new science curricula was significantly enhanced where mixed samples of male and female students were studied compared to either predominantly male or female samples.

Finally, the quantitative synthesis revealed that student performance in new science programs was adversely affected when teachers received inservice or preservice training in the use of the new curriculum materials. Alternative explanations for this and other findings are thoroughly discussed in this study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

THE EFFECTS OF NEW SCIENCE CURRICULA ON STUDENT PERFORMANCE

INTRODUCTION

Since 1955, and particularly during the 1960's and early 1970's, elementary, junior high, and secondary school science curricula experienced considerable growth and substantial change which can be described only as "phenomenal." It is generally accepted that the launching of Sputnik I on October 4, 1957 stimulated this sudden growth and concomitant curriculum development. In an attempt to "make up lost ground" in the technological race with the Soviet Union, American scientists and educators initiated an all out effort to upgrade science curricula and science instruction. The public became very science and technology conscious during this period of time as federal agencies and private foundations provided financial support for the resulting wave of new science curricula.

New science programs emerged quickly for high school physics, chemistry, and biology. The development soon encompassed the junior high and elementary science programs. Within 15 years of the historical launching of the Russian satellite, dozens of "alphabet-soup" science curricula were developed including such well-known programs as PSSC, CBA, BSCS, CHEM Study, ESS, S-APA, SCIS, and ESCP; as well as other lesser known programs such as COPES, ISLI and IS. One noted scientist and educator, Morris Shamos, estimated that approximately 5 billion dollars were spent on K-12 science improvement during the 15 year post-Sputnick era (Yager, 1981a).

A complete set of goals and objectives for the new science curricula were never really articulated by the numerous new curriculum designers. The prevailing notion, however, was that the traditional courses which tended to concentrate on the knowledge of scientific facts, laws, theories, and technological applications were somehow ineffective in developing the creative genius needed to forge ahead in a rapidly evolving scienctific world (Haney, 1966; Klopfer, 1971; Schwab, 1963). The new science curricula were supposed to rekindle student interest in science and accelerate the development of a 21st century science perspective by emphasizing the structure and process of science. Rather than allowing the students to get bogged down in the rhetoric of conclusions as was the pattern with traditional courses, the new curricula were to stress doing science and learning how to learn. New science curricula quickly came to be associated with process objectives and skills while traditional science curricula were tabbed as being fact-oriented. The process versus product characterizations of new and traditional science curricula still persists.

After nearly 25 years and over 5 billion dollars, the question of how effective new science curricula actually were in enhancing student performance is still unanswered. Money for continued development of science programs has been withdrawn and public sentiment apparently favors a move back to the basics. This move back to the basics would imply support for more traditional, fact-oriented science courses. However, the decisions to withdraw support for curriculum development and implementation and to re-emphasize traditional course objectives should be based on a careful examination of evidence, not on some gravity-like force that moves the curriculum pendulum back and forth. This report addresses the question of new science curriculum effectiveness by meta-analyzing the results of many studies which have addressed this issue during the past 25 years.

STATEMENT OF THE PROBLEM

This study was designed to synthesis quantitatively the collective research dealing with the effects of new science curricula on student performance. This meta-analysis incorporates large numbers of studies pertaining to the overall assessment and evaluation of new science curricula (versus traditional courses) on student performance. The meta-analysis approach to research integration, developed by Glass (1976), applies the attitude of data analysis to quantitative summaries of individual studies. Meta-analysis is a statistical analysis of the results of a large number of analyses of original research on a common topic.

For the purpose of this report, new science curricula are defined as those courses or curricular projects which:

a) were developed after 1955 (with either private or public funds),

b) emphasize the nature, structure, and processes of science,

c) integrate laboratory activities in daily class routine, and

d) emphasize higher cognitive skills and appreciation of science.

Traditional curricula are defined as those courses or programs which:

a) were developed or patterned after a program developed prior to 1955,

b) emphasize knowledge of scientific facts, laws, theories, and applications, and

c) use laboratory activities as verification exercises or as secondary applications of concepts previously covered in class.

In applying the above criteria to research studies reviewed, the identification of new curricula was much more clear cut than the identification of the traditional courses due to the lack of detailed information supplied in the studies. Similarly, it was difficult to establish the level of treatment fidelity in most studies for both the new curricula and the traditional; new curricula may have been used in traditional ways in some cases and vice versa. Where information about such anomalies was available, such information was coded and analyzed separately.

In addressing the overall question of the effectiveness of new science curricula developed since 1955, the data in this report are organized in three broad categories: curriculum characteristics, student or teacher characteristics, and study or design characteristics. The variable analyzed in all cases is the effect size (labeled E.S. or Δ in the remainder of this report). The effect size is a common metric derived from the criterion variable data reported in the individual studies included in the report. Representing both a magnitude and direction of group differences, the effect size metric facilitates a quantitative synthesis of individual studies in which student performance in new science programs and traditional science courses are compared.

In this meta-analysis, effect sizes were calculated for one or more of the 18 *a priori* discrete criterion variables selected for analysis. Calculated effect sizes were analyzed for each of the eighteen criteria separately and in clusters of related criteria. The individual criteria (lettered) and the criterion clusters (numbered) are as follows:

1. Achievement Cluster
    a. Cognitive - low
    b. Cognitive - high
    c. Cognitive - mixed/general achievement
2. Perceptions Cluster
    d. Affective - attitude toward subject
    e. Affective - attitude toward science
    f. Affective - attitude toward procedure/methodology
    g. Self-concept

3. Process Skills Cluster
    h. Process skills
    i. Methods of science
4. Analytic Skills Cluster
    j. Critical thinking
    k. Problem solving
5. Related Skills Cluster
    l. Reading
    m. Mathematics
    n. Social studies
    o. Communications skills
6. Miscellaneous
    p. Creativity
    q. Logical thinking (Piagetian)
    r. Spatial relations (Piagetian)

Using the effect sizes calculated from the eighteen individual criterion measures and the composite effect sizes calculated for the six criterion clusters as the dependent variables and the three broad factors (i.e., curriculum characteristics, student or teacher characteristics, and study or design characteristics) as the independent variables, a series of specific questions dealing with the effect of new science curricula on student performance were generated and analyzed. The individual criteria and criterion cluster effect size measures were analyzed by specific curriculum (e.g., PSSC, CBA, ESS), by curriculum type (e.g., physical science, life science, earth science), by grade level, community type, student gender, student race, student socio-economic status, teacher training, teacher characteristics, length of study, validity of study, curriculum profile, method of testing, and form of publication. The "Results" section of this report provides a complete description of each question analyzed along with the appropriate statistical summaries.

BACKGROUND

Science courses have been a part of the school curriculum for well over 200 years. During this period of time educational philosophies, functions, purposes, goals, and objectives have changed dramatically. Similarly, the role of science in education has changed. In assessing the impact of the new science curricula developed during the past 25 years, it seems appropriate to reflect upon some of the historical events leading up to the curriculum development era immediately following the launching of Sputnik.

To begin, what is referred to as "traditional" science courses actually are courses or textbooks written in the post-World War II era (1945-1956). Immediately following the war, the science curriculum lacked articulation and coordination. General science was considered a junior high subject, biology was typically required in 10th grade, and chemistry and physics were offered as 11th and 12th grade electives and viewed as college preparatory courses. By 1950, some additional courses such as applied science, physiology, electricity, earth science, and physical science were offered.

During the post-World War II period, less emphasis was placed on the memorization of information and more emphasis was placed on the functional aspects of science. Although information acquisition was still considered the most important goal in education, the understanding of scientific principles and the development of problem solving skills were also stressed. The laboratory gained new acceptance and importance during this period as well (Collette, 1973).

The latter stages of this period saw several changes in the science curriculum: curricula were developed for gifted science students; new courses in earth science were developed; attempts were made to correlate science with other curricular areas; and, for the first time, attention was given to elementary school science. Competition in the textbook industry intensified resulting in improved and updated textbooks and laboratory manuals in all science areas. Manufacturers of scientific materials and equipment also began making serious efforts to improve classroom products (Lacey, 1966; Richardson, 1964; Thurber and Collette, 1968).

By the mid-1950's, scientists and educators were becoming increasingly concerned over the decreasing percentage of high school students enrolled in science courses -- especially in physics. Colleges were also beginning to express concern about the quality of the student's high school science preparation (Novak, 1969; Washton, 1967). The rapid scientific and technological advances of this period began to pose a serious societal and educational problem. An understanding of science and technology was becoming imperative. As Hurd (1961) noted, the nature of this education had not yet evolved.

The National Science Foundation (NSF), conceived in the 1940's and born in 1950, was ready to go to work when the nation became concerned about the scientific capability and the status of science education in the United States. NSF began locating brilliant investigators and got them to work doing imaginative fundamental research. As NSF began to establish itself as the primary supporter of basic research, they also began to get involved with public education. The initial educational efforts were conservative in that NSF provided graduate fellowships to the brightest young scientists in order to attract them into becoming research scientists. Within a short period of time, however, NSF realized that if a dramatic growth in the scientific and technological workforce was to be accomplished without reducing quality, then the entire talent pool from which scientists are drawn had to be greatly enlarged. NSF began to support the efforts of outstanding university scientists, educators, and learning theorists in an effort to develop science courses new in conception, design, and content and to educate teachers. Many scientists turned their attention from the laboratory to the classroom and became actively involved in the curriculum reform movement. What followed was what many have come to regard as the Golden Age of Science Education (Rutherford, 1980).

One of the first tasks in order to initiate the reform was to examine the existing courses of study in science. Upon this examination of the science textbooks of the 1950's, it was evident that sporadic attempts had been made in order to keep texts up-to-date by adding bits and pieces to already existing content. The major problem, however, was that traditional topics were never deleted. Science textbooks, in general, contained a mass of often unrelated information much of which was incorrect, outdated, and irrelevant to modern science. The conclusion was that existing courses were not able to be salvaged and that new courses of study in line with modern science and modern learning theory would have to be developed (Collette, 1973).

The curriculum reform movement had a gradual beginning with the formal organization of the Physical Science Study Committee (PSSC) late in 1956. This committee was the result of the 1954 recommendation of the Division of Physical Science of the National Academy of Science that professional physicists work with high school and college instructors in order to develop new physics courses and materials. The plan was to bring about "immediate" change.

The result was that some of the most innovative and spectacular changes ever to occur in American public school education took place in the area of science (Collette, 1973). The public became very science and technology conscious. Along with the increased public support came increased financial support from federal agencies and private foundations. From 1956-1967 the NSF contribution to curriculum reform projects at all levels exceeded $100,000,000 (Welch, 1968). NSF also substantially increased the number of programs to improve the science backgrounds of teachers. Colleges and universities established institute programs which offered courses in science and mathematics in order to update teachers. Whereas in 1953, there were only two NSF summer institutes in science and mathematics, in 1963 there were 412 such institutes with about 21,000 teachers receiving instruction (Science Policy Research Division Report, 1975).

By 1970, after a decate and a half of curriculum development and implementation, the United States had apparently established a preeminence in science education to match its status in basic scientific research (Rutherford, 1980). The hundreds of millions of dollars spent on curriculum development and implementation generally was felt to be a good investment (Conant, 1976; Schlessinger and Helgeson, 1969; Welch, 1968). Unfortunately though, many people felt that the job had been accomplished, and thus the nationally funded curriculum efforts began to slow down rapidly. A small cadre of science educators claimed that only part of the job had been completed and urged NSF to continue its work in the area of science education (Rutherford, 1980).

During the period of curriculum development, implementation, and in-service institutes numerous evaluation studies were completed to assess the impact of these innovative programs on student performance. The most typical assessment was a comparative study measuring one or more student outcome variables with one of the new curricula as a treatment group and a traditional science course as a control group. By the mid 1970's, however, curriculum assessment and evaluation efforts began to taper off without any real conclusive evidence that the Golden Age of Science Education had produced any substantial gains besides updating the subject matter.

During the 1975-76 academic year teacher education activities were suspended. In 1976, NSF responded to Congressional pressure and awarded contracts to assess the current status of science education at the elementary and secondary levels (Butts, *et al.*, 1980; Yager, 1981a).

NSF funded a Status Study of three major independent but related studies to be conducted in parallel (Rutherford, 1980). Each study was designed from a different perspective to assess the status of science education in the United States (Helgeson, Blosser, and Howe, 1978; Stake and Easley, 1978; Weiss, 1978).

The focus of the Helgeson, Blosser and Howe (1978) Status Study, conducted at the Center for Science and Mathematics Education, The Ohio State University, was to report on the impact of activity in curriculum development, teacher education, instruction, and needs in science education. Specifically, the purpose of their study was to:

> 1. review, analyze and summarize the appropriate literature related to pre-college science instruction, to science teacher education, and to needs assessment efforts; and
> 2. identify trends and patterns in the preparation of science teachers, teaching practices, curriculum materials, and needs assessments in science education during the period, 1955-1975. (Helgeson, Blosser, and Howe, 1978, p. 1)

Their report is divided into five major sections. One section deals with existing practices and procedures in schools, another summarizes science teacher education, the following section deals with controlling and financing education, the next reports on needs assessment efforts, and the final section presents a summary and trends of needs and practices.

The second Status Study was organized by a team of researchers at the University of Illinois and was co-directed by Stake and Easley (1978). Case Studies in Science Education is a collection of field observations of science teaching and learning in American public schools during the school year 1976-77. The study was undertaken to provide NSF with a portrayal of the current conditions in K-12 science classrooms to help make NSF's programs of support for science education consistent with national goals and needs. Eleven high schools and their feeder schools were selected to provide a diverse and balanced group of sites. Field researchers were on-site from 4 to 15 weeks and were instructed to find out what was happening and what was felt to be important in science (including methematics and social science) programs. Each observer prepared an in-depth case study report which was presented intact as part of a final collection and later augmented with cross-site conclusions by the Illinois team.

The third Status Study was directed by Weiss (1978) of the Research Triangle Institute. The purpose was to design and implement a national survey to answer the following questions:

1. What science courses are currently offered in schools?*
2. What local and state guidelines exist for the specification of minimal science experiences for students?
3. What texts, laboratory manuals, curriculum kits, modules, etc., are being used in science classrooms?
4. What share of the market is held by specific textbooks at the various grade levels and subject areas?
5. What regional patterns of curriculum usage are evident? What patterns exist with respect to urban, suburban, rural, and other geographic variables?
6. What "hands-on" materials, such as laboratory or activity centered materials, are being used? What is the extent and frequency of their use by grade level and subject matter?

---

* The National Science Foundation defines science to include the natural sciences, social sciences, and mathematics.

7. What audio-visual materials (films, filmstrips/loops, models) are used? What is the extent, frequency and nature of their use by grade level and subject area?

8. By grade level, how much time (in comparison with other subjects) is spent on teaching science?

9. What is the role of the science teacher in working with students? How has this role changed in the past 15 years? What commonalities exist in the teaching styles/strategies/practices of science teachers throughout the United States,

10. What are the roles of science supervisory specialists at the local district and state levels? How are they selected? What are their qualifications?

11. How have science teachers throughout the United States been influenced in their use of materials by Federally-supported in-service training efforts in science? (Weiss, 1978, p. 1)

This survey utilized a national probability sample of districts, schools, and teachers. The sample was designed so that national estimates of curriculum usage, course offerings and enrollments, and classroom practices could be made from the sample data. The sample included superintendents, supervisors, principals, teachers, and other school personnel.

The Office of Education (OE) also funded a project to assess the status of science education. The third assessment of science as part of the National Assessment of Educational Progress (1978) provides information regarding the results of science instruction in the United States. This report is a comprehensive assessment of science knowledge, skills, attitudes, and educational experiences of precollege students (Kahl and Harms, 1981). The third assessment included a new battery of items which provided information regarding affective outcomes of science education for nine-, thirteen-, and seventeen-year-olds, as well as for an adult sample (Yager, in press).

In 1978, NSF funded a project to synthesize and to interpret the information from the three K-12 Status Study reports and the NAEP assessment. This research effort, called "Project Synthesis", examined K-12 science education from five perspectives (biology, physical science, inquiry, elementary school science, and science/technology and society) within four goal clusters and critical elements for teaching (e.g., instructional procedures, teacher characteristics, instructional facilities and materials, and others) (Yager, 1981a).

In an attempt to increase the scope of the three K-12 Status Study reports, NSF (1980) selected nine professional organizations with different responsibilities and perspectives to analyze the studies independently and submit reports. The organizations selected were:

Teacher Organizations
1. National Council for the Social Studies
2. National Council of Teachers of Mathematics
3. National Science Teachers Association

Science Organizations
1. American Association for the Advancement of Science
2. National Academy of Science

Administration and Support Organizations
1. American Association of School Administrators
2. Association for Supervision and Curriculum Development
3. National Congress of Parents and Teachers
4. National School Boards Association

(Rutherford, 1980)

The reports of these organizations provide an interesting and informative view regarding the totality of science education in American schools (Rutherford, 1980) and are available in the NSF document entitled, "What Are the Needs in Precollege Science, Mathematics, and Social Studies Education? Views from the Field."

Finally, in the spring of 1979, NSF funded a Status Study of Graduate Science Education in the United States, 1960-1980 (Yager, 1980b). The purpose of this project was to consider the current status of science education at graduate institutions. This study was viewed as an extension of the three Status Studies for K-12 science education (Helgeson, Blosser, and Howe, 1978; Stake and Easley, 1978; Weiss, 1978) and as a logical next step to consider the unique features of the discipline of science education as perceived by science educators from institutions throughout the United States. Funds from this project also allowed a summer writing group to assemble at The University of Iowa. A paper entitled "Crisis in Science Education" resulted from this effort (Yager, 1980a).

The three K-12 Status Study reports (Helgeson, Blosser, and Howe, 1978; Stake and Easley, 1978; Weiss, 1978), the professional reviews of the Status Study reports (NSF, 1980), and the reports proclaiming a crisis in science education (Yager, 1980a, 1980b) all provide an interesting assessment of where science education has been, where science education is today, and where science education should be headed. But the hard truth is that none of the reports has stimulated the interest of public or private groups to the extent that the groups conducting the studies had originally hoped. The qualitative nature of these assessments may explain the diminished impact of the results.

Critics tend to question the overall validity of qualitative analyses where problems with investigator bias are difficult to control.

Quantitative synthesis techniques considerably reduce the potential for investigator bias. A meta-analysis of research focusing on the various criterion variables, criterion clusters, and criterion clusters by study variables provides a comprehensive assessment of the effects of the new curricula on student cognitive and affective achievement. Such a comprehensive assessment should establish specific and firm conclusions of value to practitioners.

Since curriculum revision and evaluation is a continuing process, the conclusions of this meta-analysis are important to researchers assessing curriculum development and implementation for several reasons: (1) those areas where questions of interest have already been adequately answered will be identified; (2) those areas where the research results are inconclusive or are not worthy of further investigation will be identified; (3) those questions which have not yet been adequately explored will be revealed. This knowledge should result in fewer research projects being devoted to duplication of research which does not appear to be necessary, fewer research projects being devoted to unimportant questions or issues, and more research projects being directed to the major questions which are yet unanswered. Such a synthesis is long overdue.

Finally, the results of this study have potential significance for groups which establish public and educational policies, as well as groups which implement these policies. A comprehensive assessment of the effectiveness of science curricula developed since 1955 should provide valuable information for future development and reseach activities.

PROCEDURES

This study was designed to investigate the impact and effects of the new curricular programs developed for elementary, junior high, and secondary science education since 1955. The meta-analysis perspective of research integration, developed by Glass (1976), is utilized to record quantitatively the properties and findings of studies which measured and compared student performance in a new science curricula with student performance in a traditional course. Only studies involving United States samples are included in this meta-analysis. This groundrule was established since the curricula studied were originally designed for use in American schools and generally modifications are made when these curricula are adopted for use internationally.

This section includes a description of: the research methods involved in conducting this meta-analysis, the studies included in this meta-analysis, the coding variables and coding reliability, procedures regarding effect size calculations, and methods of data analysis.

Description of the Search Methods

The first task in conducting a meta-analysis is to locate and obtain the relevant research studies in the field of interest. The first step in this project was to collect and examine a representative sample of science education research studies in order to map out the research literature to be meta-analyzed. Literature was sampled across time and type of publication from the following sources: Dissertation Abstracts International, The Journal of Research in Science Teaching, Science Education, and the most recent abstracts of presentations for the National Association for Research in Science Teaching annual convention. Literature searches were then conducted on a sampling basis to obtain an estimate of the size of the relevant literature. These searches were conducted using Dissertation Abstracts International, ERIC, Social Science Research, and Science Education: A Dissertation Bibliography (1978).

Arrangements were made with the ERIC Center to borrow the large number of dissertation microfilms which were identified. It was determined that the sequence of searching and the subsequent coding of documents would be as follows: dissertations, research documents and reports available from ERIC or on microfiche, published journal articles, and other documents identified during the coding process. The rationale for the above order was the desirability of beginning with primary and most comprehensive sources, as well as to avoid any duplication of data in situations where researchers later reported all or part of their research studies in professional journals. The final stage of the search procedure was to review the following journals for relevant studies reported from 1955 to 1980 which were not previously coded: American Biology Teacher, High School Journal, Journal of Chemical Education, Journal of Research in Science Teaching, Journal of Secondary Education, The Physics Teacher, Science Education, The Science Teacher, and School Science and Mathematics.

Studies Included

Three hundred two studies were examined for this meta-analysis. One hundred five of those studies contained sufficient data for the meta-analysis. Studies included in the meta-analysis had to satisfy the following criteria:

1. Studies had to be conducted at the elementary, junior high, or secondary level between 1955 and 1980. College level studies were included if the curricula were not modified and if the students had no prior course in that science discipline.
2. Studies had to be conducted in the United States using United States samples. Thus, comparative studies between United States samples and international samples were not included.
3. Studies had to be an experimental investigation comparing student performance in a new science curricula to student performance in a traditional course (e.g., ESS versus traditional, ESCP versus traditional, BSCS versus traditional). Descriptive or theoretical studies are not included in this meta-analysis, nor are studies which only reported student performance on variables for which there was no control group.

Three hundred forty-one effect sizes (Δ's) were calculated from the studies included in the meta-analysis. Table 1 shows the distribution of effect sizes by source of study. These studies represent a

TABLE 1

DISTRIBUTION OF EFFECT SIZES (Δ's) BY SOURCE OF STUDY

Effect Sizes per Study	SOURCE					
	Dissertations		ERIC Documents		Journal Articles	
	Studies	Δ's	Studies	Δ's	Studies	Δ's
1	20	20	2	2	6	6
2	20	40	4	8	7	14
3	9	27	1	3	3	9
4	6	24	0	0	5	20
5	4	20	1	5	0	0
6	6	36	0	0	1	6
7	2	14	1	7	0	0
8	2	16	0	0	1	8
9	0	0	0	0	0	0
10	1	10	1	10	0	0
11	0	0	0	0	0	0
12	0	0	0	0	0	0
13	0	0	0	0	0	0
14	0	0	0	0	0	0
15	0	0	0	0	0	0
16	1	16	0	0	0	0
17	0	0	0	0	0	0
18	0	0	0	0	0	0
19	0	0	0	0	0	0
20	1	20	0	0	0	0
TOTAL	72	243	10	35	23	63

total sample size of 45,626 students. Table 2 shows the distribution of the student sample by source of study.

TABLE 2

DISTRIBUTION OF STUDENT SAMPLE
BY SOURCE OF STUDY

	Treatment	Control	Total
Dissertations (N = 72)	13,987	14,569	28,556
ERIC Documents (N = 10)	4,145	3,462	7,607
Journal Articles (N = 23)	4,645	4,818	9,463

Coding Variables

There are numerous study characteristics which can influence the effectiveness of treatments in comparative studies. A critical part of this meta-analysis involved identifying and coding factors related to studies. In order to make full use of statistical methods in the meta-analysis, various features of each study were measured or otherwise expressed in quantitative terms. Many of these features are expressed in familiar scales (e.g., date of publication, length of study in weeks, IQ of students, grade level) while other features are nonordinal characteristics which are coded by indicator variables (e.g., form of publication, secondary school background, curriculum profile, rated internal validity, the specific characteristics of the treatment). The coding form utilized in this study was developed by the investigating team during a week-long meeting and meta-analysis training session in Boulder, Colorado. (Refer to Appendix A for the complete coding form.) The coding form is subdivided into the following categories: background and coding information, sample characteristics, treatment characteristics, teacher characteristics, and effect size calculation. Each of these categories is discussed below.

Each study was read and a coding form was completed for each outcome and each comparison in the study. A list of coding conventions was developed during the week-long training session. These were used to guide the classification of studies whose characteristics were ambiguous. These conventions are also explained below.

Background and Coding Information

The numeric coding of each study extended across two computer cards -- 176 digits of coding in all. The *reader ID#* identified the number of the card in the data record of two cards (i.e., ID#: 1 or 2). Each study was identified by a reader code and a study code. The reader code identified the project site and the researcher at that site who coded the study.

The *comparison code* refers to the number of different treatments compared to a control group within a study. A comparison code of 01 01 would indicate one comparison within the study while a comparison code of 02 03 would indicate the second comparison or treatment group of a total of three treatments (e.g., ESS, SCIS and S-APA all compared to a control group).

The *outcome code* refers to the number of dependent outcome variables assessed in the study. The coding system of outcomes for a study is the same as comparisons within a study. Thus, an outcome code of 01 03 refers to the first identified variable coded of a total of three variables coded (e.g., a single study may have assessed cognitive factors, affective factors, and critical thinking).

The *date of publication* was recorded as stated on the coded manuscript. In some cases studies were published more than once. In these cases the most complete source was coded. If the manuscripts were similar, then the earliest date of publication was recorded.

The *form of publication* was classified according to the form in which the coded study appeared: journal article, book, MA/MS thesis, dissertation, or unpublished manuscript. The most complete source of data was recorded. Thus, if a dissertation was later published in a journal, the dissertation was coded.

Sample Characteristics

A number of variables were coded which were specifically related to the student sample of each study included in the meta-analysis.

The *grade level* of the students was coded and classified into five categories: primary (K-3), intermediate (4-6), junior high (7-9), senior high (10-12), and post secondary. The post secondary classification was included for any studies which might have used one of the newly developed curricula at the community college or college level.

The *total sample size* represents the total number of students in the treatment and control groups.

The *length of the study* was coded in weeks indicating the duration of the treatment. Sequential studies were coded up to a duration of three years. All sequential studies longer than three years duration were categorized together.

Gender was coded as the percentage of female students in each study. For studies which did not state the percentage of males and females this figure was inferred. For elementary, junior high, and required secondary science courses the percentage of female students in the study was inferred to be 50%. For chemistry courses the percentage of female students was coded as 25%. For physics courses the percentage of female students was inferred to be in the 10-15% range depending upon the total sample size. This range was used for rounding purposes since physics studies generally had fewer subjects.

The *average ability* of the students was recorded on the basis of low (below 95 IQ), average (95-105 IQ), or high (above 105 IQ). The *homogeneity of the IQ* was recorded as homogeneous or heterogenous, as well as the *source of IQ* (i.e., whether it was stated within the study or inferred). If the average ability of the students in the sample was inferred, it was recorded as being heterogeneous average IQ if the sample was an elementary, junior high, or required secondary science course. If the sample was a chemistry course the average ability was coded as high ability heterogeneous. If the sample was a physics course the average ability was inferred to be high ability homogeneous.

The *race* of the sample and the *predominant minority* was coded if the information was provided in the study. Race was recorded as the percentage of non-white students. The predominant minority categories were: Mexican, non-Mexican Hispanic, Oriental, American Indian, Black, or other. The *percentage of predominant minority* was also recorded if that information was provided.

The *socio-economic status* of the sample was coded as low, medium, or high. The *homogeneity of the socio-economic status* was also recorded (i.e., homogeneous or heterogeneous). In some instances this information was inferred from the geographic location of the study site.

The *secondary school science background* was coded for each study. The following courses were coded either as "yes" or "no" regarding the secondary student's prior science: life science (typically a 7th grade course), physical science (typically an 8th grade course), general science or earth science (typically 9th grade courses), biology (typically a 10th grade elective), chemistry (typically an 11th grade elective), and physics (typically a 12th grade elective). If the students' science background was not stated in the study, it was inferred that they had taken all courses prior to the science course they were currently enrolled in with the exception of earth science.

The *handicapped* variable was used to code any studies in which the student sample involved any of the following physical or emotional handicaps: visually impaired, hearing impaired, learning disability, emotionally disturbed, multiple handicaps, or educable mentally retarded.

The sample size of students in both the treatment and control groups was recorded (*N of pupils in* T_1 and *N of pupils in* T_2). The

% mortality of T_1 and T_2 was also recorded if that figure was reported in the study.

The *special grouping by ability* variable was used to code whether students were grouped into a low, medium, or high track; or, whether students were not grouped by ability.

The *size of the school* involved in the study was coded as stated. The following criteria were used: less than 50 students, 50-199 students, 200-499 students, 500-999 students, 1000-1999 students, greater than 2000 students.

The *type of community* was also coded as stated in the study, or inferred on the basis of geographic location of the study site, as follows: rural, suburban, or urban.

Treatment Characteristics

The *treatment code* variable refers to the elementary, junior high, or secondary science curricula which was used as the treatment course in each study coded. The majority of these curricula were identified prior to any coding. A few curricula, however, were added to the coding list after the coding process began. A complete list of curricula treatment groups follows:

- Elementary Science
  - Elementary Science Study (ESS)
  - Science Curriculum Improvement Study (SCIS; or, SCIIS, SCISII)
  - Science - A Process Approach (S-APA)
  - Outdoor Biology Instructional Strategies (OBIS)
  - Elementary Science Learning by Investigation (ESLI)
  - ESSENCE
  - Conceptionally Oriented Program for Elementary Science (COPES)
  - Modular Activities Program in Science (MAPS)
  - Unified Science and Mathematics for Elementary Schools (USMES)

- Minnesota Mathematics and Science Teaching Project (MINNEMAST)
- Individualized Science (IS)
- Science Curriculum for Individualized Learning (SCIL)
- Elementary School Training Program in Scientific Inquiry (University of Illinois) (ESTPSI)
- Flint Hills Elementary Science Project (Kansas State Teachers College) (FHESP)

Junior High Science
- Human Science Program (HSP)
- Time, Space and Matter (TSM)
- Individualized Science Instructional System (ISIS)
- Intermediate Science Curriculum Study (ISCS)
- Introductory Physical Science (IPS)
- Earth Science Curriculum Project (ESCP)
- Interaction of Matter and Energy (IME)
- Conservation Education/Environmental Education/Ecology (CE/EE)
- Montclair Science Project (MSP)

Secondary Science
- Biological Science Curriculum Study (BSCS)
  - Special Materials (BSCS/SM)
  - Yellow Version (BSCS/Y)
  - Blue Version (BSCS/B)
  - Green Version (BSCS/G)
  - Advanced Materials (BSCS/A)
- Chemical Education Materials Study (CHEM Study)
- Chemical Bond Approach (CBA)
- Physical Science Study Committee (PSSC)
- Harvard Project Physics (HPP)
- Conservation Education/Environmental Education/Ecology (CE/EE)
- Physical Science for Nonscience Students (PSNS)
- Interdisciplinary Approaches to Chemistry (IAC)

A *curriculum profile* was established for the major elementary, junior high, and secondary science curricula. The profile assessed each curriculum on five parameters: (1) degree of inquiry, (2) emphasis on process skills, (3) emphasis on the laboratory and/or laboratory skills, (4) degree of individualization, and (5) emphasis on content. Each parameter was ranked from low (1) to high (4). The scores for each curricula represent an average score based on assessments of five science educators familiar with each of the programs. The curriculum profile of major curricula was developed for two purposes: (1) to be able to record any modifications made within the context of each individual study regarding any of the parameters, and (2) to make comparisons between curricula. The *study modification to curriculum profile* variable indicates whether modifications were made toward the low end of each curriculum profile category, toward the high end of each curriculum profile category, or whether there were no modifications made. See Table 3 for curriculum profile data.

The *technology used* variable indicates whether hand held calculators, films, television, or computers were used or not within the study.

Teacher Characteristics

For studies which stated the ratio of male to female teachers involved in the experiment, the *percentage of female teachers* was recorded. If reported, the *average number of years of science teaching experience* was coded; as well as, the *average number of years teaching science curriculum* T_1, and the *average number of years teaching science curriculum* T_2.

The *race* of the teachers involved in the study and the *predominant minority* was coded if the information was provided. The predominant minority categories were: Mexican, non-Mexican Hispanic, Oriental, American Indian, Black, and other. The *percentage of predominant minority* was also recorded if that information was provided.

The *average educational background* for teachers involved in each study was coded as follows: less than a Bachelors degree, Bachelors

TABLE 3

CURRICULUM PROFILE

	Inquiry	Process Skills	Emphasis on Lab	Degree of Individualization	Emphasis on Content
Elementary Curricula					
ESS	4	3	4	4	1
SCIS	3	3	3	3	2
S-APA	2	4	3	2	3
OBIS	3	2	3	2	2
ESLI	2	2	2	2	2
ESSENCE	1	1	1	4	1
COPES	2	3	2	2	3
MAPS	2	3	3	2	3
USMES	3	3	3	2	1
MINNEMAST	2	2	3	2	3
Junior High Curricula					
TSM	2	2	2	2	3
ISIS	3	4	3	3	2
ISCS	2	2	4	3	4
IPS	2	3	4	2	2
ESCP	2	2	3	2	4
IMP	2	2	3	2	3
Secondary Curricula					
BSCS (Special Materials)	3	3	4	4	3
BSCS Yellow	2	3	3	2	3
BSCS Blue	2	3	2	2	4

(continued)

TABLE 3 (Continued)

	Inquiry	Process Skills	Emphasis on Lab	Degree of Individualization	Emphasis on Content
Secondary Curricula (continued)					
BSCS Green	3	3	3	2	3
BSCS Advanced	3	3	4	4	3
CHEM Study	2	3	3	2	3
CBA	1	2	2	1	4
PSSC	1	3	3	2	4
Project Physics	2	3	3	3	3

degree, Bachelors plus 15 hours, Masters degree, Masters plus 15 hours, Masters plus 30 hours, Doctorate degree.

The remaining coding variables in this section deal with teacher training: *was preservice training provided?*; and, *was inservice training provided?*. The financial funding of inservice training was coded if such information was provided: locally funded and/or sponsored; university funded and/or sponsored; federally funded.

Design Characteristics

A characteristic often considered important in judging the quality of a comparative study is how the experimenter allocated subjects to the treatment and control groups. The *assignment of students to groups* variable represents whether students were randomly assigned to groups, selected in matched pairs, part of intact groups, or volunteered to be a part of the experiment (self-selecting). The *assignment of teachers to groups* variable was coded for random assignments, non-random assignments, self-selecting assignments, or for situations where teachers taught both groups (crossed) or were matched on certain measures.

The *unit of analysis* variable coded whether individual students, a classroom of students, an entire school, or some other group of students was used as the primary unit of analysis in the study.

The *type of study* was coded according to Campbell and Stanley (1963) definitions as: correlational, quasi-experimental, experimental, or pre-experimental.

The *rated internal validity* was judged on the basis of the assignment of subjects to groups and the extent of subject mortality in the study. Low internal-validity studies were those whose matching procedures were weak or nonexistent, or where intact convenience samples were used. The study was also rated low if mortality was exceptionally high or severely disproportionate. Medium internal validity ratings were assigned according to the following criteria: (1) studies with randomization but high or differential mortality, or (2) studies with "failed" randomization procedures (e.g., where the experimenter began by randomizing, but then resorted to other allocation methods) and low mortality, or (3) studies with intact groups but highly similar and low mortality, or (4) extremely well-designed matching studies. To be judged high on the internal validity measure, a study must have used random assignment of subjects to groups and have low and fairly equivalent mortality rates.

Occassionally, statistical or measurement irregularities decreased the level of internal validity (e.g., when an otherwise well-designed study employed different testing times for the treatment and control groups). It is also recognized that other factors such as sample size, congruence of the measures with the treatment or control groups, the method of measurement, or the reactivity of the measurement influence internal validity. These five constructs were assessed separately.

Outcome Characteristics

The *content of measure* variable identified the science discipline involved in the study: life science, physical science, general science, earth science, biology, chemistry, or physics. All elementary studies were coded as general science.

The *congruence of measure with* T_1 and T_2 is a measure of test reactivity. Congruence was measured as low, medium, or high. For example, if a general achievement test designed specifically for PSSC was used to compare achievement of PSSC students versus non-PSSC students, the congruence for T_1 (treatment group) was coded high and the congruence for T_2 (control group) was coded low.

The *type of criterion* refers to the twenty-two criterion variables identified for coding. The eighteen variables for which data were obtained were grouped into six criterion clusters for analysis. The six criterion clusters and the eighteen individual criterion variables are listed in the "problem statement" section.

The *criterion measured* variable identifies whether the study assessed student performance or teacher performance. There were no studies included in this meta-analysis which assessed teacher performance.

The *method of measurement* indicates whether the study measurement was: a standardized test; an *ad hoc* written test (e.g., developed by researcher, curriculum project); observational (e.g., passive or instructional observations); or, a structural interview or assessment.

The *reactivity* of the measurement refers to the level of researcher bias in the tests used. Standardized tests were considered to have low reactivity while experimenter-made tests were judged to have high reactivity.

Effect Size Calculations

The *source of effect size data* variable refers to whether the effect size was: calculated directly from reported data or raw data from the study (e.g., means and variances); reported with direct estimates (e.g., ANOVA, t-test, F-values); calculated directly from frequences reported on ordinate scales (Probit, X^2); calculated backwards from variance of means with randomly assigned groups; calculated from nonparametric statistics (other than X^2); guessed from independent sources (e.g., test numbers, other students using the same test, conventional wisdom); estimated from variance of gain scores (correlational guessing); or, derived from probability level only (i.e., conservative estimates).

The *source of means* was coded as reported in each study. The following categories were used: unadjusted post-test; covariance adjusted; residual gains; pre-post-test differences; or, other.

The *reported significance* of each study was coded as: $p \leq .005$; $.005 < p \leq .01$; $.01 < p \leq .05$; $.05 < p \leq .10$; or, $p > .10$.

The *dependent variable units* were coded if they were reported in grade-equivalent units or some other unit. The *mean difference in grade-equivalent units* was reported if the dependent variable was reported in grade-equivalent units.

If the group variances were observed individually, then the *ratio of experimental to control group variances* was calculated, as well as the *effect size based on experimental group variance (A)*, the *effect size based on control group variance (B)*, and an *average effect size based on (A) and (B)*. If the group variances were not observed individually, the *study effect size* was reported directly from the source of the effect size data.

Reliability

Once the coding variables were identified and ground rules were established, estimates of coder reliability were calculated. The reliability of a measurement "is the statement which represents the various sources of error in the repeated measurement of a single phenomenon or the consistency in which an individual performs the same task over a period of time" (Brown and Webb, 1968, p. 37). An instrument itself is neither reliable nor unreliable -- it is only when the instrument has been used to collect data that one can speak sensibly about reliability.

Based upon a random sampling of five studies read and coded independently by the two coders involved in the study, a 94.8% coder agreement was attained in coding the 76-80 study variables (studies which reported group variances individually contained 80 study variables while studies which did not report group variances individually had only 76 study variables).

Procedures Regarding Effect Size Calculation

The magnitude of the effect of a treatment is the most important variable in any outcome study. In this study, the effect of new curricula on student performance was assessed by measuring the magnitude and direction of change for twenty-two criterion variables. Meta-analysis involves calculating a common metric for defined variables within a study. The common metric, measuring the magnitude of the effect, is referred to as an effect size (abbreviated E.S. and symbolized by the Greek letter Δ). The effect size is a normalized measure of the performance difference of two groups on a dependent variable (e.g., general achievement, critical thinking, self-concept). Effect size is defined as the mean difference between treatment conditions divided by within-group standard deviation (Glass, 1976).

$$\text{E.S.} = \frac{\overline{X}_t - \overline{X}_c}{SD_c}$$

where: $\overline{X}_t$ = mean of treatment group;

$\overline{X}_c$ = mean of control group; and

SD_c = standard deviation of control group.

Nearly all of the effect sizes calculated for this study used either the formula above or, in studies which reported *F*-values, the

F-value was considered equal to t^2 and the following formula was used:

$$\text{E.S.} = t\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where: t = t-value;

n_1 = sample size of treatment group; and

n_2 = sample size of control group.

If only the total sample size (N) was reported, it was assumed that $n_1 = n_2$ since equal n's provide a more conservative estimate of the effect size than unequal n's.

In a few instances, the only information reported in the study was that a particular test statistic (e.g., t or F or Fisher's Z - transformation of N) was calculated on *n cases* with a level of significance p. Provided that the p-value is reported exactly and not rounded, the transformation is straight forward. If, for example, it is reported that a two group t-test with $n_1 = n_2 = 6$ is significant at the $p = .02$ level (two-tailed test), then it is a simple matter of determining the corresponding t-value:

$$_{.99}t_{10} = 2.76.$$

Knowing n_1, n_2, and the value of the t-test, the effect size can be calculated using the formula:

$$\Delta = t\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= 2.76\sqrt{\frac{1}{6} + \frac{1}{6}}$$

$$= 1.59.$$

In studies reporting only an approximation of the p-value for a measured criterion variable, the conservative value of p was utilized in order to estimate the value of t. This yields a conservative effect size.

The reader is referred to Glass, McGaw, White and Smith (1980, p. 136-197) for detailed derivations and illustrations of procedures for transforming other reported statistics and measurement scores into effect sizes.

Methods of Data Analysis

During the coding phase of this meta-analysis a total of three hundred two studies were reviewed. One hundred five of those studies contained sufficient data for the meta-analysis. From these studies, which repr ent a total sample size of 45,626 students, three hundred forty-one effect sizes (E.S.) were calculated. The coding form for this meta-analysis (Appendix A) provides information regarding the variables for which data were collected (i.e., background information, sample characteristics, treatment characteristics, teacher characteristics, design characteristics, outcome characteristics, and effect size calculations).

Thus, in response to the overall question assessing the effects of new curricular programs developed in science education since 1955, summary statistics of effect sizes were calculated for: the 27 new science curricula for which data were collected; the 18 student performance measures; and the 6 criterion cluster variables. Sample characteristics such as grade level, community type, length of study, student gender, and socio-economic status; background information regarding form of study publication; treatment characteristics; content characteristics; internal validity; the curriculum profile characteristics; in-service training of treatment instructors data; and, method of measurement of the criterion variable data were also coded. The relationship between the six criterion cluster measures and each of the subsequent variables listed above was analyzed.

Effect size summary statistics calculated for each of the variables listed above were: mean effect sizes, minimum Δ, maximum Δ, standard deviation, and t-values. Statistical analysis was accomplished using the General Linear Model (GLM) procedure of the *Statistical Analysis System* (SAS) (Helwig and Council, 1979) on the IBM 370/168 at The University of Iowa (programs were run under release 79.4B). The model source statement of the GLM gives the dependent variables and independent effects. Due to unequal cell frequencies orthogonality is destroyed. A condition for orthogonality is that the number of observations in each combination of treatments is equivalent (Hayes, 1973). Thus, the Type IV Sum of Squares (SS) is used as described in SAS-76 (Barr *et al.*, 1976). The corrected total reported for each analysis is equal to the number of effect sizes in the data set minus one (N-1). The Duncan's Multiple Range Test for specified effect size variables was also calculated and these data are reported where appropriate.

All statistically significant data in this report are identified by an asterisk (*). Such values are significant at the *a priori* alpha level of 0.05.

RESULTS AND DISCUSSION

Curriculum Characteristics

How do students exposed to various new science curricula compare to students exposed to traditional science curricula on a composite performance level?

The literature search revealed 105 codable studies comparing new science curricula to traditional science programs in terms of one or more student performance criteria. The codable studies encompassed 27 different science curricula and 18 distinct performance criteria. As an overall indicator of new curriculum effectiveness, effect size data extracted from all studies on a specific new science curricula were summarized. This summary of a composite performance analysis is presented in Table 4 for the 27 new curricula included in this meta-analysis. Although the data in Table 4 do not provide information about the specific focus of the original research studies analyzed, the composite student performance data by new curricula do provide a starting place -- a first approximation regarding the effectiveness of new science curricula.

TABLE 4

EFFECT SIZE DATA FOR COMPOSITE STUDENT PERFORMANCE MEASURES* BY CURRICULUM

CURRICULUM	N	MEAN Δ	MINIMUM Δ	MAXIMUM Δ	S.D.	t-value
Elementary						
ESS	11	0.37	0.01	0.81	0.27	4.56*
SCIS	45	0.30	-0.82	2.41	0.55	3.64*
SAPA	45	0.27	-0.57	2.50	0.55	3.26*
USMES	17	0.55	-0.01	4.48	1.05	2.17*
MINNEMAST	2	1.51	0.55	2.47	1.35	1.57
IS	2	0.64	0.60	0.69	0.06	14.33*
SCIL	16	0.43	-0.32	1.12	0.38	4.44*
ESTPSI	6	0.39	0.07	0.73	0.28	3.35*
FHESP	1	-0.06	---	---	0.00	---
Junior High						
HSP	4	0.66	0.46	0.85	0.18	7.24*
TSM	1	0.49	---	---	0.00	---
ISCS	6	0.18	-0.10	0.74	0.31	1.39
IPS	17	0.00	-0.44	0.44	0.28	0.04
ESCP	24	0.16	-0.70	0.86	0.38	2.11*
IME	4	0.22	-0.33	0.66	0.43	1.02
CE/EE	1	0.01	---	---	0.00	---
MSP	3	0.21	0.11	0.42	0.17	2.13
Secondary						
BSCS/SM	4	0.11	0.01	0.29	0.12	1.72
BSCS/Y	28	0.48	-0.50	1.78	0.57	4.36*
BSCS/B	6	2.32	0.44	4.18	1.41	4.03*
BSCS/G	5	0.13	-0.18	0.34	0.21	1.36
BSCS/A	4	0.09	-0.17	0.43	0.29	0.62

continued

TABLE 4 (Continued)

CURRICULUM	N	MEAN Δ	MINI-MUM Δ	MAXI-MUM Δ	S.D.	t-value
CHEM	33	0.12	-0.49	0.92	0.37	1.84
CBA	16	0.24	-0.81	1.09	0.45	2.15*
PSSC	35	0.47	-1.04	2.70	0.69	4.08*
HPP	2	0.28	0.27	0.30	0.02	19.00*
PSNS	3	0.11	0.08	0.15	0.03	5.28*

* Value is significant at the *a priori* alpha level (α.05).

Thus, Table 4 contains a summary of the number of effect sizes calculated for each curricula (N), the mean effect size, the maximum and minimum effect size, and the standard deviation of the effect size around the mean. Also listed is the t-value for the test of statistical difference between the mean effect size calculated and zero.

Recall that by definition the effect size is a measure of the mean differences in performance between students in new science curricula and students in traditional courses divided by the within group standard deviation of the control group (traditional course). Thus, an effect size of zero indicates that there were no observable differences between the two groups for the composite performance measures. A positive effect size signifies that students in the treatment group (new curricular group) performed better than the control group for the observed measures of student performance; whereas, a negative effect size signifies that student scores in the control group (traditional course) were higher.

The composite data in Table 4 clearly indicate that students who were exposed to new science curricula performed better than their traditional course counterparts. Disregarding curricula where only one effect size was calculated (FHESP, TSM, and junior high conservation education/environmental education) the average composite student performance measure effect sizes range from $\Delta = 0.00$ for IPS to $\Delta = 2.32$ for BSCS (Blue Version). It should also be noted that the most heavily studied curricula ($N > 15$) also show a definite positive impact. The average composite student performance effect sizes for these curricula range from $\Delta = 0.00$ for IPS to $\Delta = 0.55$ for USMES with 80% of the curricula in this category being statistically significant from zero at the *a priori* alpha level of 0.05. Furthermore, if these average effect sizes are translated into percentile scores, the average USMES student performed better than 71% of the traditional course students whereas the average BSCS (Blue Version) student performed better than 99% of their traditional course counterparts.

How do students exposed to new science curricula compare to students exposed to traditional science curricula on various performance criteria?

Another general indicator of new curriculum effectiveness is provided in the breakdown of effect size data for each of the 18 performance criteria measured. An examination of the mean effect sizes in Table 5 indicates that the new curricula had a positive impact on student performance for every performance criteria except for student

TABLE 5

EFFECT SIZE DATA FOR PERFORMANCE CRITERIA

CRITERION	N	MEAN Δ	MINI-MUM Δ	MAXI-MUM Δ	S.D.	t-value
Achievement:						
Cognitive-Low	8	0.02	-0.46	0.50	0.29	0.20
Cognitive-High	11	0.05	-0.49	0.41	0.28	0.60
Cognitive-Mixed	111	0.43	-1.04	4.18	0.77	5.83*
Perceptions:						
Affective-Subject	6	0.51	0.00	0.85	0.32	3.87*
Affective-Science	25	0.50	0.11	1.75	0.36	6.89*
Affective-Method	10	0.41	-0-81	1.20	0.58	2.25*
Self-Concept	10	-0.08	-0.82	0.82	0.53	- 0.51
Process Skills:						
Techniques	28	0.61	-0.10	2.50	0.66	4.83*
Methods of Science	28	0.17	-0.62	0.73	0.32	2.79*
Analytical Skills:						
Critical Thinking	31	0.19	-0.36	1.44	0.37	2.77*
Problem Solving	4	0.71	0.06	1.41	0.70	2.02
Related Skills:						
Reading	23	0.10	-0.41	0.92	0.24	1.99*
Mathematics	18	0.40	-0.50	4.48	1.07	1.59
Social Studies	2	0.25	0.25	0.26	0.00	51.00*
Communications	5	0.40	0.08	0.75	0.26	3.47*
Miscellaneous:						
Creativity	5	0.71	0.18	1.50	0.50	3.22*
Spatial Relations	2	0.57	0.29	0.86	0.40	2.02

*Value is significant at the *a priori* alpha level (0.05).

self-concept. Eleven of these positive differences were found to be statistically significant from zero.

The small number of effect sizes available for some criteria may limit a meaningful interpretation of those criteria (e.g., spatial relations and social studies). However, the consistent pattern of positive effect size values clearly establishes the superiority of the new science curricula over traditional courses in enhancing student performance over a broad range of performance measures.

Especially interesting in the composite data of Table 5 are the statistics for general achievement (N = 111). Much criticism regarding the new science curricula focused on the apparent decline of general science knowledge among students exposed to the new programs. At the height of the new curricular movement (and even today) the prevailing notion was that the process goals of the new science curricula were being achieved at the expense of the content goals -- although no comprehensive data base existed for either claim. The data in Table 5 show clearly that students exposed to new science curricula achieved 0.43 standard deviations above (exceeding 67% of the control group) or, nearly one-half of a grade level better than, their traditional curriculum counterparts.

In the areas where most new curriculum opponents would concede superiority, Table 5 indicates consistently positive effect size patterns. Student attitudes toward the subject specifically, science generally, and the new format of the courses (method) all show statistically significant positive results. Similarly, the areas involving higher cognitive skills (e.g., problem solving, critical thinking, logical thinking, and creativity) show consistently positive effect size patterns. Even student performance in related areas such as reading, mathematics, and communication skills, areas in which new curriculum proponents often purported student gains, show positive effect size data.

The slightly negative effect size mean for the self-concept data appear, at first, to be an anomaly when considered with the other affective measures regarding student attitude toward the specific subject, science, and methods. However, in the majority of the studies coded, the self-concept measures assessed global self-concept rather than subject-specific self-concept. Thus, one would not expect the global self-concept to change dramatically during a period of 21-36 weeks (the average length of treatment in the studies coded). In fact, when considering the goals and objectives of the new curricula and the emphasis upon student decision making, one might expect student self-concept to be deflated a little at the outset and duration of a course. Thus, the slightly negative, near zero effect size (Δ = -0.08) for student self-concept appears to be predictable and reasonable.

The cumulative effect size data in Table 5 make it possible to examine the impact new science curricula had on specific areas of student performance such as achievement, attitudes toward science, techniques of science, or critical thinking. However, the small number of effect sizes focusing on certain individual performance parameters (e.g., problem solving, N = 4; spatial relations, N = 2) and the obvious relationship between other performance parameters suggested the need for a smaller number of more broadly defined criterion clusters. Moreover, the larger number of effect sizes within criterion clusters facilitate the examination of more detailed questions regarding new curriculum effectiveness. The criterion clusters and the individual performance parameters comprising the clusters are listed below:

Achievement Cluster
  a. Cognitive-low (Recall of facts, laws, principles)
  b. Cognitive-high (Application, synthesis, evaluation)
  c. Cognitive-mixed (General achievement)

Perceptions Cluster
  d. Affective-attitude toward subject
  e. Affective-attitude toward science
  f. Affective-attitude toward method/class environment
  g. Affective-attitude toward self (self-concept)

Process Skills Cluster
  h. Techniques of science (lab skills, measurement)
  i. Methods of science

Analytic Skills Cluster
  j. Critical thinking
  k. Problem solving

Related Skills Cluster
  l. Reading (comprehension/readiness)
  m. Mathematics (concepts, skills, applications)
  n. Social studies (content, skills)
  o. Communication skills (reading, writing, speaking)

Miscellaneous

- p. Creativity
- q. Logical thinking (Piagetian tasks)
- r. Spatial relations (Piagetian tasks)

Table 6 contains descriptive statistics and t-test data for effect sizes grouped by criterion clusters. The graph of the effect size means for the clusters is shown in Figure 1. Consistent with the Table 5 data for individual performance criteria, the criterion clusters data indicate that students exposed to new science curricula consistently outperformed students exposed to traditional courses.

How do students exposed to specific new science curricula compare to students in traditional science courses on the six criterion cluster measures (i.e., achievement, perceptions, process skills, analytic skills, related skills, other areas)?

The analysis of effect size data for specific curricula by criterion culsters is inherently interesting because of the detail provided. The increased detail is accompanied by a decrease in available studies from which effect size data can be extracted. With 18 separate criterion variables studied across the 27 new science curricula coded in this study, the full matrix would require 486 effect size calculations to place a minimum of one effect size in each cell. Even with the clustering of effect size data across related dependent variables, many of the possible cells yielded no data.

TABLE 6

EFFECT SIZE DATA FOR CRITERION CLUSTERS ACROSS ALL CURRICULA

VARIABLE	N	MEAN Δ	MINI-MUM Δ	MAXI-MUM Δ	S.D.	t-value
Achievement	130	0.37	-1.04	4.18	0.73	5.76*
Perceptions	51	0.37	-0.82	1.75	0.49	5.40*
Process Skills	56	0.39	-0.62	2.50	0.56	5.17*
Analytic Skills	35	0.25	-0.36	1.44	0.44	3.29*
Related Skills	48	0.25	-0.50	4.48	0.69	2.51*
Other Areas	21	0.33	-0.70	4.50	0.51	2.93*

*Value is significant at the *a priori* alpha level (0.05).

Achievement Cluster

Table 7 lists the effect size data for the achievement criterion cluster for the 20 new science curricula for which such data were available. The mean effect size was positive for all but two of the curricula (FHESP, N = 1, Δ = -0.06 and TME, N = 2, Δ = -0.11) indicating that students in the new science programs overwhelmingly outperformed students in traditional courses on achievement measures. The effect size results are especially impressive for students enrolled in the BSCS-Yellow (N = 19, Δ = 0.45), PSSC (N = 23, Δ = 0.51) and SCIS (N = 5, Δ = 1.00) programs. The skeptic might dismiss these results due to inherent sampling problems caused by students gravitating to the new science programs such as BSCS or PSSC. Thus, the

FIGURE 1

BAR GRAPH OF THE MEAN EFFECT SIZES FOR CRITERION CLUSTERS ACROSS ALL CURRICULA

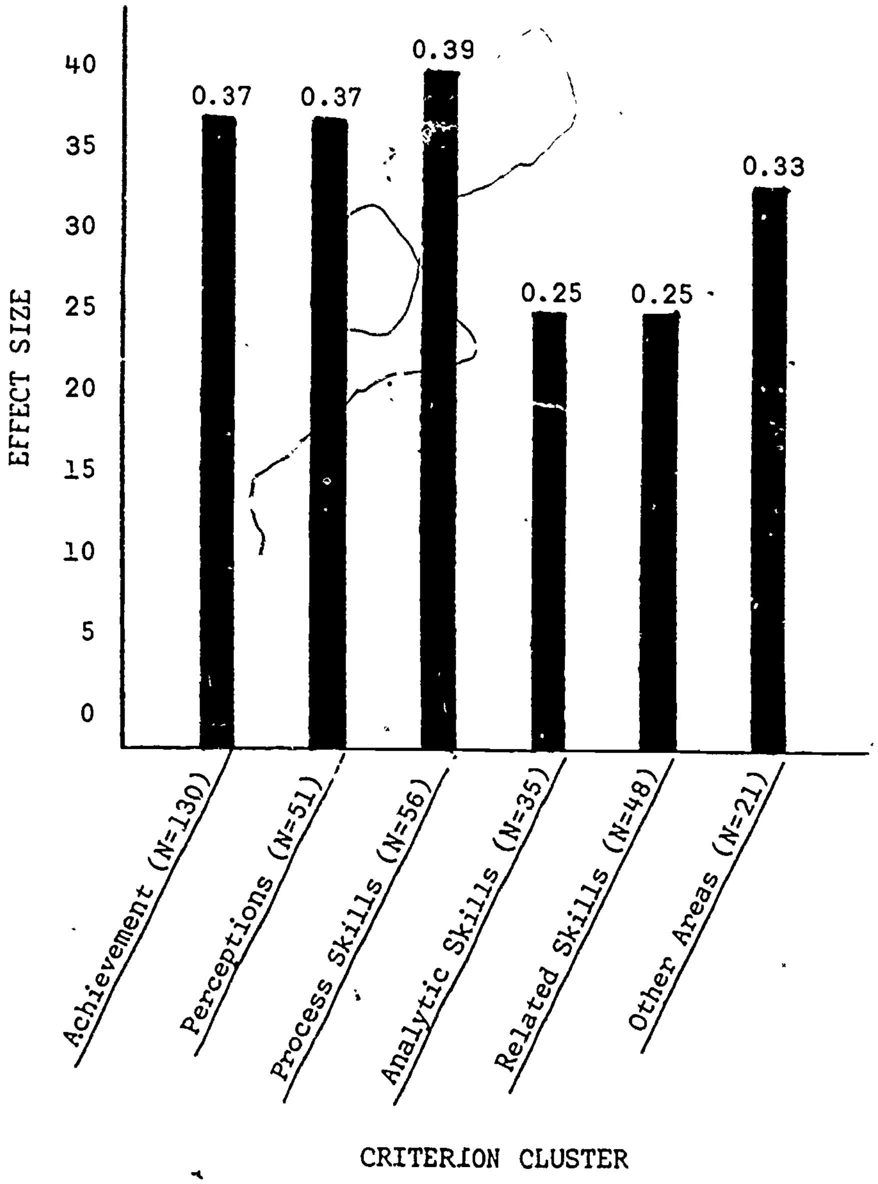


TABLE 7

EFFECT SIZE DATA FOR THE ACHIEVEMENT CRITERION CLUSTER BY CURRICULUM

CURRICULUM	N	MEAN Δ	MINIMUM Δ	MAXIMUM Δ	S.D.	t-value
ESS	3	0.09	0.01	0.24	0.12	1.34
SCIS	5	1.00	0.05	2.41	0.91	2.44
SAPA	12	0.17	-0.57	1.65	0.58	1.04
USMES	3	0.34	0.11	0.54	0.21	2.74
MINNEMAST	2	1.51	0.55	2.47	1.35	1.57
SCIL	4	0.06	-0.03	0.20	0.11	1.18
ESTPSI	3	0.28	0.07	0.60	0.27	1.79
FHESP	1	-0.06	---	---	0.00	---
IPS	3	0.03	-0.27	0.20	0.26	0.24
ESCP	6	0.19	-0.52	0.86	0.49	0.97
IME	2	-0.11	-0.33	0.10	0.30	- 0.53
MSP	1	0.42	---	---	0.00	---
BSCS/SM	2	0.02	0.01	0.03	0.01	2.00
BSCS/Y	19	0.45	-0.49	1.78	0.54	3.67*
BSCS/B	2	3.94	3.70	4.18	0.33	16.42*
BSCS/G	2	0.17	0.00	0.34	0.24	1.00
BSCS/A	4	0.09	-0.17	0.43	0.28	0.62
CHEM	23	0.12	-0.49	0.92	0.40	1.37
CBA	10	0.27	-0.42	1.09	0.41	2.05
PSSC	23	0.51	-1.04	2.70	0.77	3.16*

*Value is significant at the *a priori* alpha level (0.05).

potential for a sampling error owing to superior students self-selecting a new, innovative science program must be considered as a threat to the internal validity of new curriculum studies. The question of self-selection bias is addressed separately in a later analysis (see Table 28). An examination of the data in Table 7 across all curricula would suggest however, that self-selection errors in the original studies were either not a factor, or they tended to produce inconsistent effects; since most of the effect size means are not statistically significant from zero.

Focusing on some of the more common new science curricula, the data in Table 7 indicate substantial gains in achievement for students in: SCIS ($\Delta = 1.00$), USMES ($\Delta = 0.34$), BSCS-Blue ($\Delta = 3.94$, PSSC ($\Delta = 0.51$), BSCS-Yellow ($\Delta = 0.45$), and CBA ($\Delta = 0.27$). In all of these curricula, the average student in the new science curricula exceeds the achievement scores of 60% of the students in the control group ($\Delta = 0.25$ is equivalent to the 60th percentile of the control group). Furthermore, other major curricula with moderate effects such as IPS ($\Delta = 0.03$), ESS ($\Delta = 0.09$), CHEM Study ($\Delta = 0.12$), BSCS-Green ($\Delta = 0.17$), S-APA ($\Delta = 0.17$) and ESCP ($\Delta = 0.19$), equivalent to a percentile ranging from the 51st to the 58th percentile of the control group, should be viewed quite favorably as support for the philosophy and effects of the new science curricula since the major objections of these programs centered around their lack of emphasis regarding science content. Clearly, the data in Table 7 indicates that students enrolled in new science programs are not stifled in their acquisition of scientific knowledge.

Student Performance

The data in Table 8 relate to the comparison of students in new science curricula versus students in traditional courses on attitudes toward the specific subject matter, the broad area of science, the classroom climate, and the students themselves (self-concept). The analysis reveals significantly enhanced student attitudes in 5 of the 9 new curricula where multiple effect sizes were coded. Of the more popular curricula, the S-APA ($\Delta = 0.39$) and HSP ($\Delta = 0.66$) showed the most positive effects while SCIS ($\Delta = 0.08$) and CBA ($\Delta = 0.16$) showed the least positive changes.

All of the curricular means for the perceptions cluster reveal positive effects, in spite of the fact that negative effects were coded for SCIS, SCIL and CBA. Thus, in terms of effective measures students generally felt better about the specific course they were taking, the methods employed, science in general, and themselves while enrolled in a new science program. It should be noted however, that only 16 of the 27 curricula studied had been investigated within this large domain called perceptions (over 40% of these curricula had only 1 effect size calculated). However, data in Table 20 would tend to indicate that student perceptions by grade level were greatly enhanced, i.e., elementary grades ($\Delta = 0.28$), junior high ($\Delta = 0.59$) and high school ($\Delta = 0.44$).

Process Skills

Process objectives have become synonymous with new science curricula over the years. The debate over the relative importance of

TABLE 8

EFFECT SIZE DATA FOR THE PERCEPTIONS CRITERION CLUSTER BY CURRICULUM

CURRICULUM	N	MEAN Δ	MINIMUM Δ	MAXIMUM Δ	S.D.	t-value
ESS	1	0.51	---	---	0.00	---
SCIS	14	0.08	-0.82	0.82	0.52	0.59
SAPA	6	0.39	0.00	0.76	0.28	3.35*
USMES	1	0.15	---	---	0.00	---
IS	2	0.64	0.60	0.69	0.06	14.33*
SCIL	8	0.61	-0.32	1.12	0.44	3.86*
HSP	4	0.66	0.46	0.85	0.18	7.24*
ISCS	1	0.17	---	---	0.00	---
IPS	2	0.23	0.21	0.25	0.02	11.50*
ESCP	1	0.11	---	---	0.00	---
IME	2	0.55	0.45	0.66	0.14	5.29
BSCS/Y	1	1.05	---	---	0.00	---
BSCS/G	1	1.75	---	---	0.00	---
BSCS/B	2	0.25	0.21	0.29	0.05	6.25
CBA	4	0.16	-0.81	0.76	0.69	0.45
PSNS	1	0.15	---	---	0.00	---

*Value is significant at the *a priori* alpha level (0.05).

process skill development versus content knowledge acquisition drew considerable attention in the sixties and early seventies. The issue still stimulates discussion today even though many teachers have resigned themselves to a content emphasis. The data in Table 9 deal with the process skill record of the new science curricula. The research record clearly indicates a success story for most of the new

TABLE 9

EFFECT SIZE DATA FOR THE PROCESS SKILLS CRITERION CLUSTER BY CURRICULUM

CURRICULUM	N	MEAN Δ	MINIMUM Δ	MAXIMUM Δ	S.D.	t-value
ESS	4	0.47	0.26	0.70	0.18	5.19*
SCIS	6	0.56	0.12	1.20	0.36	3.72*
SAPA	3	1.08	-0.02	2.50	1.28	1.46
USMES	1	0.29	---	---	0.00	---
SCIL	4	0.43	0.38	0.47	0.04	18.34*
ESTPSI	3	0.50	0.15	0.73	0.30	2.81
ISCS	1	0.30	---	---	0.00	---
IPS	5	-0.08	-0.44	0.23	0.30	- 0.63
ESCP	8	0.22	-0.62	0.52	0.39	1.60
BSCS/SM	1	0.11	---	---	0.00	---
BSCS/Y	4	0.72	0.23	1.76	0.71	2.01
BSCS/B	1	2.45	---	---	0.00	---
CHEM	5	-0.03	-0.33	0.26	0.22	- 0.37
CBA	1	0.34	---	---	0.00	---
PSSC	5	0.35	-0.10	1.19	0.49	1.60
HPP	2	0.28	0.27	0.30	0.02	19.00*
PSNS	2	0.09	0.08	0.10	0.01	9.00

*Value is significant at the *a priori* alpha level (0.05).

curricula. But not all curricula show equal success. Both the IPS (Δ = -0.08) and CHEM Study (Δ = -0.03) accumulated a record of negative performance on process skills development.

The slightly negative results for both the IPS and CHEM Study are especially interesting because of the emphasis both place on the integration of laboratory activities in total course work. The lack of success shown by the studies of the curricula suggest ineffective curricular materials or improper implementation of potentially effective materials. Again, the lack of detailed information describing the treatment conditions (i.e., with how much integrity were the new materials implemented) prohibits a thorough investigation of the alternatives.

The remaining data in Table 9 follow a trend similar to the achievement data in Table 7 where effect size data are available on both criteria for a particular curriculum. The curricula showing strong positive effect size values in achievement show similar values in process skill measures (e.g., SCIS $\Delta_A = 1.00$, $\Delta_{P.S.} = 0.56$; BSCS-Yellow $\Delta_A = 0.65$, $\Delta_{P.S.} = 0.72$; PSSC $\Delta_A = 0.51$, $\Delta_{P.S.} = 0.35$).

Analytic Thinking

Perhaps the area most stressed by new science curriculum developers in the golden years of curriculum reform was problem solving and critical thinking. Capturing just the right mixture of text material, laboratory activity, and stimulating problems was the dream of every curriculum engineer and the challenge awaiting every new science curriculum teacher. Were the new materials being developed in the post-Sputnik years any more effective in cultivating student analytic thinking skills than the traditional courses they were replacing? Many new science curriculum appeared and disappeared before that question was fully explored. In fact, it is questionable that the issue was really explored fully considering Table 10 indicates only 35

TABLE 10

EFFECT SIZE DATA FOR THE ANALYTIC THINKING CRITERION CLUSTER BY CURRICULUM

CURRICULUM	N	MEAN Δ	MINI-MUM Δ	MAXI-MUM Δ	S.D.	t-value
SAPA	1	0.06	---	---	0.00	---
ISCS	1	0.07	---	---	0.00	---
IPS	5	-0.15	-0.36	0.12	0.22	- 1.57
ESCP	7	0.16	-0.05	0.44	0.18	2.31*
CE/ES	1	0.01	---	---	0.00	---
MSP	1	0.12	---	---	0.00	---
BSCS/SM	1	0.79	---	---	0.00	---
BSCS/Y	3	0.42	0.03	1.08	0.57	1.27
BSCS/B	2	0.94	0.44	1.44	0.70	1.88
BSCS/G	1	-0.18	---	---	0.00	---
CHEM	5	0.30	-0.08	0.75	0.32	2.08
CBA	1	0.21	---	---	0.00	---
PSSC	6	0.53	0.01	1.41	0.61	2.12

*Value is significant at the *a priori* alpha level (0.05).

codable effects addressing the question of new science curriculum impact on student analytic thinking skill with only 5 curricula revealing more than one effect size.

The four most frequently studies curricula (IPS, ESCP, CHEM, and PSSC) show slightly mixed results. The IPS studies showed an overall negative impact while the other three showed a positive effect with PSSC being the highest at 0.53 standard deviations. Perhaps the most surprising data on analytic thinking are those for the BSCS curricula. The 7 studies on these curricula produced a mean effect size of $\Delta = 0.46$ second only to physics ($\Delta = 0.53$). For a subject area generally considered non-quantitative at the high school level, these results are very impressive.

Related Skills

The related skills cluster contains those studies conducted to determine the effects of new science curricula on mathematics skills, reading skills, social studies performance, and communications skills (e.g., writing and speaking). The promise of enhanced student performance in related skill areas was never advertised loudly by new curriculum proponents, but the inference that gains in these areas could be achieved as an added benefit was able to be concluded from much of the early rhetoric.

As Table 11 indicates, only three of the new science curricula actually were studied to any extent for their impact on related skill areas: SCIS, SAPA, and USMES, all elementary level programs. While the mean of the USMES study effect sizes is the most impressive ($\Delta = 0.66$), the top-end value of 4.48 and the resulting standard deviation for the effect size data of 1.25 leaves the overall mean somewhat suspect. It is probably safe to conclude however, that student performance in related skill areas was positively enhanced through their participation in such curricula.

TABLE 11

EFFECT SIZE DATA FOR THE RELATED SKILLS CRITERION CLUSTER BY CURRICULUM

CURRICULUM	N	MEAN Δ	MINIMUM Δ	MAXIMUM Δ	S.D.	t-value
SCIS	13	0.21	-0.09	0.54	0.15	5.00*
SAPA	18	0.10	-0.41	0.75	0.29	1.46
USMES	12	0.66	-0.01	4.48	1.25	1.84
ISCS	2	-0.03	-0.10	0.03	0.09	- 0.54
MSP	1	0.11	---	---	0.00	---
BSCS/Y	1	-0.50	---	---	0.00	---
PSSC	1	0.04	---	---	0.00	---

*Value is significant at the *a priori* alpha level (0.05).

Other Performance Areas

As indicated by the relatively small number of effect size values listed in Table 12, the number of studies focusing on non-conventional measures of student performance is relatively small. Fourteen of the effect sizes included in the table are derived from studies using Piagetian-type tasks and 5 of the studies utilized creativity measures as a dependent variable. The paucity of studies on these variables for any one curriculum makes meaningful synthesis difficult. Perhaps the most significant conclusion to be drawn from these data is that more experimental studies need to be conducted using these criteria as dependent variables.

TABLE 12

EFFECT SIZE DATA FOR THE OTHER MENTAL FUNCTIONS CRITERION CLUSTER BY CURRICULUM

CURRICULUM	N	MEAN Δ	MINI-MUM Δ	MAXI-MUM Δ	S.D.	t-value
ESS	3	0.48	0.06	0.81	0.38	2.18
SCIS	7	0.18	-0.55	0.86	0.55	0.88
SAPA	5	0.50	-0.09	1.50	0.62	1.82
TSM	1	0.49	---	---	0.00	---
ISCS	1	0.74	---	---	0.00	---
IPS	2	0.35	0.26	0.44	0.12	3.89
ESCP	2	-0.13	-0.70	0.44	0.80	- 0.23

How do students exposed to new science curricula of a particular content area (physics, chemistry, biology, earth science, etc.) compare to students in traditional science courses?

Effect size data are grouped by the science content area for the analysis reported in this section. For each study reviewed and coded, the content area represented by the curriculum under study was placed in one of the following categories:

(1) Life science

(2) Physical science

(3) General science

(4) Earth science

(5) Biology

(6) Chemistry

(7) Physics

As was mentioned earlier, the ground rule in coding dictated that the study be labeled with the highest degree of specificity rather than the most general. Thus, PSSC physics or CBA chemistry are classified as physics and chemistry, respectively, not physical science. Elementary science curricula were coded as general science.

Life Sciences

This grouping of science curricula includes those dealing with topics such as health science and junior high life science.

As indicated in the data in Table 13 the student perceptions criterion cluster is the only one for which multiple effect sizes were calculated. The strong positive effect size mean ($\Delta = 0.66$) indicates that the students exposed to the new life science programs developed more positive attitudes about science than students participating in the standard health and life science programs.

Physical Science

The curricula included in the physical science category represent junior high school programs for the most part. Because of the recent interest in the junior high/adolescent student expressed by funding agencies, these data and the subsequent grade level analyses found later in this report are significant. The two criterion clusters showing the most dramatic differences among the effect size data reported for the physical science curricula in Table 13 are achievement and perceptions. The combination of these two sets of performance data suggest that the new curricula represented by these junior high studies had a positive impact on the student participants. The only negative impact is found in the area of problem solving/critical thinking where a slightly negative effect size appears ($\Delta = -0.10$).

General Science

Just as the physical science curricula are most characteristic of junior high programs, the general science curricula are comprised mostly of the elementary school programs. Perhaps the most revealing statistic of the Table 13 data is that there are 143 effect sizes included under the general science category -- almost half of all the effect sizes calculated from the codable studies. The relative wealth of research in this content area is most likely a function of the numbers of students enrolled in elementary science programs compared to those enrolled in upper grade level programs and the accessability of elementary school populations.

If a consistency of performance of new curricula in any content area is sought, the general science area is far and away the winner. In all 5 performance areas where multiple studies were located, the effect size data indicate that students participating in the new programs performed significantly better than their traditional course counterparts. The performance of the average elementary student in the new science curricula coded exceeds 61-72% ($\Delta = 0.27$ to $\Delta = 0.59$) of the students in traditional science courses for these 5 criterion clusters.

Earth Science

Earth science curricula tend to be used at the 9th grade level although some of the new earth science programs have been used as high school electives and as advanced 7th/8th grade courses. As the data in Table 13 indicate, studies performed in the earth science area are the only ones that produced statistically significant differences in the analytic skills criterion cluster (the high school science areas, biology, chemistry, and physics each produced substantial positive effect sizes, but not statistically significant from zero).

Contrasted to the significant results produced in the area of analytic skills, the earth science curricula also distinguish themselves as the only content area for which a positive achievement result was not achieved (Δ = -0.07); however, this mean is not significantly different from zero.

Biology

New science curricula in biology are synonymous with BSCS. The collapsed category of biology represents a composite view of research completed on the various versions of BSCS (Special Materials, Yellow, Blue, Green, and Advanced). Of the high school science programs developed, more codable research was found for BSCS programs than any other single project.

An examination of the data in Table 13 reveals an impressive track record regarding the research results on the BSCS programs. Where multiple effect sizes of a performance cluster exist, the mean effect size values are consistently high ($\Delta \geq 0.46$). One of the more interesting positive effects of BSCS is in the area of analytic thinking. The studies coded yielded an effect size mean of 0.46. This mean is higher than that generated for all chemistry curriculum research reviewed ($\Delta = 0.28$) and approaches the mean of the physics program research ($\Delta = 0.53$). Considering that traditional biology is noted for its preoccupation with facts and labels, the mean of 0.46 is quite an impressive turn-around.

Chemistry

Two chemistry curricula comprised the market of new curricula during the decade of the sixties: CHEM Study and CBA. Of the three traditional high school subject areas (biology, chemistry, and physics), it is probably safe to conclude on the basis of data in Table 13 that the new chemistry curricula produced the least impact in terms of enhanced student performance. The mean effect size of 0.16 for the studies on achievement, while statistically significant from zero, is not as impresive as Biology ($\Delta = 0.59$) or Physics ($\Delta = 0.50$). Only the achievement data for the earth science curricula ($\Delta = -0.07$) yielded a smaller mean effect size.

An even less impressive figure for the new chemistry programs is in the area of process skills. Recall that process skill measures are those reflecting an understanding of and familiarity with laboratory procedures, designing, executing, and interpreting experiments, and problem solving procedures that involve active participation. One would expect this to be a forte of the new chemistry programs. This does not appear to be the case. While the mean effect size of 0.28

in the area of analytic skills is respectable, it does not make up for the dismal record in the process skill area.

Physics

Studies of two physics curricula, PSSC and HPP were coded for this meta-analysis. The data in Table 13 include 35 effect sizes generated from PSSC studies and 2 from HPP studies. Interestingly, no codable studies were found which dealt with student perceptions. Apparently, researchers were interested more in the cognitive performance areas than in the affective domain. Yet, a grave concern still exists over the declining enrollments in science and especially in high school physics.

The performance chart of the new physics programs is second only to the biology curricula in the overall positive effect size pattern. Studies of achievement and analytic skills yielded mean effect sizes of about a half standard deviation. Translated into grade equivalents, this means students participating in the new physics courses effectively gained a half-year of study on their traditional course classmates in terms of general physics achievement and analytic thinking skills. The decline, perhaps demise is a better work, of the new physics programs should not be attributed to a lackluster performance based on the research reviewed!

TABLE 13

EFFECT SIZE DATA FOR PERFORMANCE CRITERION CLUSTERS BY CONTENT OF CURRICULA

CLUSTER	N	MEAN Δ	MINI-MUM Δ	MAXI-MUM Δ	S.D.	t-value
Life Science						
Perceptions	4	0.66	0.46	0.85	0.18	7.24*
Analytic Skills	1	0.01	---	---	0.00	---
Physical Science						
Achievement	9	0.31	-0.33	1.14	0.47	1.97
Perceptions	8	0.31	0.15	0.66	0.16	5.48*
Process Skills	10	0.08	-0.44	0.78	0.34	0.74
Analytic Skills	7	-0.10	-0.36	0.12	0.21	- 1.32
Other Areas	6	0.27	-0.70	0.74	0.50	1.35
General Science						
Achievement	32	0.35	-0.57	2.47	0.68	2.95*
Perceptions	30	0.32	-0.82	1.12	0.49	3.55*
Process Skills	19	0.59	0.12	2.50	0.52	4.95*
Analytic Skills	1	0.06	---	---	0.00	---
Related Skills	46	0.27	-0.41	4.48	0.69	2.67*
Other Areas	15	0.35	-0.55	1.50	0.53	2.53*
Earth Science						
Achievement	4	-0.07	-0.52	0.27	0.32	- 0.45
Perceptions	1	0.11	---	---	0.00	---
Process Skills	8	0.22	-0.62	0.52	0.39	1.60
Analytic Skills	7	0.16	-0.05	0.44	0.18	2.31*
Biology						
Achievement	29	0.59	-0.49	4.18	1.04	3.07*
Perceptions	4	0.82	0.21	1.75	0.72	2.28

(continued)

TABLE 13 (Continued)

CLUSTER	N	MEAN Δ	MINIMUM Δ	MAXIMUM Δ	S.D.	t-value
Process Skills	6	0.90	0.11	2.45	0.96	2.29
Analytic Skills	7	0.46	-0.18	1.44	0.58	2.09
Related Skills	1	-0.50	---	---	0.00	---
Chemistry						
Achievement	33	0.16	-0.49	1.09	0.40	2.28*
Perceptions	4	0.15	-0.81	0.76	0.69	0.45
Process Skills	6	0.02	-0.33	0.34	0.25	0.24
Analytic Skills	6	0.28	-0.08	0.75	0.29	2.40
Physics						
Achievement	23	0.50	-1.04	2.70	0.77	3.16*
Process Skills	7	0.33	-0.10	1.19	0.40	2.18
Analytic Skills	6	0.53	0.01	1.41	0.61	2.12
Related Skills	1	0.04	---	---	0.00	---

* Value is significant at the *a priori* alpha level (0.05).

Compared to students in traditional science courses, how is student performance affected by the level of emphasis on inquiry, process skills, laboratory, individualization, and content across new science curricula?

Studies focusing on the effectiveness of a particular science program such as PSSC, CHEM, or BSCS dominate the research literature on new science curricula. The difficulties of doing large scale research across curricular or content lines explains the abundance of the focused studies. However, there is interest in questions that cut across program and content lines. Questions such as, "how does the amount of emphasis on inquiry affect student performance?" are of interest. But they defy easy investigation because of the large samples needed to override the interactive effects of any one program or science area. This is not practical in original research. The quantitative synthesis of study results permits a *post hoc* analysis of such questions even though the original studies may not have focused on the issue.

Using the ratings of a panel of five science educators, profiles of the science curricula encountered in the research studies analyzed were constructed. Each curriculum was rated on a scale of 1 (low) to 4 (high) on the level of emphasis on: (1) inquiry, (2) process skills, (3) laboratory, (4) individualization, and (5) content. With the available information, profiles were constructed for 21 of the 27 curricula encountered. The profiled curricula accounted for 306 of the 341 effect sizes available for analysis.

Effect size data were grouped as "high" (3 or 4) or "low" (1 or 2) on each of the five profile factors and analyzed for each performance criterion cluster. The effect size data for each separate profile factor are listed in Tables 14-18.

Inquiry Emphasis

Since the explanation of the curriculum profile data constitutes an exhaustive report by itself, a brief presentation of some of the major features of the analyses are presented here. Perhaps the first point of interest lies in the profile ratings themselves. The panel evaluating the various curricula rated 73% of the new curricula low on inquiry, 80% high on process skills, 93% high on laboratory emphasis, 78% low on individualization, and 73% high on content emphasis. Certainly, these ratings reflect the bias of the panel and represent a source of error in interpreting the results. Especially interesting are the low ratings on inquiry and individualization -- ratings that run counter to the original goals purported by curriculum architects and assumed by uninformed teachers and lay people.

The data on the inquiry factor in Table 14 do not reveal any overall pattern of performance but do show data of interest in the perceptions and related skills areas. With an even distribution of high and low rated curricula (N = 18), the 0.42 effect size associated with curricula with a low rating on inquiry appears to be considerably more positive than that associated with curricula rated high on inquiry. Assuming traditional curricula would receive a low inquiry rating by these same panelists, these data suggest that the positive affective student response to new science programs is not a function of the inquiry nature of the new materials, but of some other factor(s). This relationship was explored more fully by correlating the curriculum profile ratings with the effect size data across all performance measures and by each performance cluster. As indicated by the correlation data reported in Table 19, student perception effect size data consistently correlate negatively to the four factors most often attributed to new science curricula (inquiry, process orientation, laboratory emphasis, and individualization).

The related skills data are just the opposite of the student perception data when analyzed by level of emphasis on inquiry. Effect size values are considerably higher ($\Delta = 0.42$ versus $\Delta = 0.06$, Table 14) for studies conducted on curricula with a high inquiry rating. The mean data for effect size grouped by emphasis on Content (Table 18) and the correlation data in Table 19 substantiate a firm pattern which shows that student performance in related skill areas is positively affected in science curricula which emphasize inquiry and negatively affected in curricula which stress content.

Process Emphasis

The effect size analysis on the curriculum profile rating process skills (Table 15) has two interesting features in addition to the predictable finding that student performance on process measures is enhanced considerably in curricula which stress process skills ($\Delta = 0.50$ (high) versus $\Delta = 0.12$ (low)). The analysis indicates that student performance on analytic skill and related skill measures is increased significantly when the curricula are rated high on process skill emphasis ($\Delta = 0.38$ versus $\Delta = 0.06$ for analytic skills; and, $\Delta = 0.27$ versus $\Delta = 0.01$ for related skills). The strong positive correlation between analytic skill performance and curriculum process skill profile is borne out in Table 19 as well ($r = 0.34$). This strong process orientation-analytic skill performance relationship is one that many proponents of the new curriculum movement purported, that many opponents doubted, and that few studies ever addressed.

Laboratory Emphasis

The role of the laboratory in school science is an issue that generates emotional debate at all levels of science education: Is it or is it not a critical component of instruction? The effect size data analyzed by the degree of emphasis placed on the laboratory in new science curricula are presented in Tables 16 and 19. An examination of the means in Table 16 reveals a definite pattern: in the four performance areas where a dichotomy of high and low emphasis on laboratory could be established, students participating in studies where the curriculum under investigation was rated "low" on laboratory emphasis consistently outperformed students participating in studies where the curriculum involved was rated "high." Notice, however, that the means for the low laboratory emphasis curricula were based on relatively few effect sizes compared to the high emphasis and that none of the low emphasis means were significantly different from zero. Recall that only 7% of the effect sizes (N = 21) were calculated from studies involving curricula receiving a low rating on laboratory emphasis. The correlation data in Table 19 further suggests that the increased laboratory emphasis of the new science curricula is not necessarily a case of "more is better." Indeed, the negative correlation between laboratory emphasis ratings and effect size data on student perceptions criteria (r = -0.34) suggest that students aren't as positive about new science program experiences when the emphasis on laboratories is increased.

Individualization Emphasis

Table 17 contains the results of effect size data analyzed by the degree of individualization of the new curricula. As evidenced by the data, achievement results are slightly enhanced in curricula judged higher on individualization while analytic skills are not. True to form, student perceptions appear to be adversely affected by increased individualization (Δ = 0.40 (low) versus Δ = 0.11 (high)). This observation is supported by the r = -0.39 correlation coefficient in Table 19.

Content Emphasis

Table 18 contains the effect size data analyzed by the level of emphasis on content. The data reveal substantial disparities in only two areas: student perceptions (Δ = 0.42 (high) versus Δ = 0.12 (low)) and related areas (Δ = 0.42 (low) versus Δ = 0.06 (high)). However, the slightly more positive effect size for the low content emphasis curricula on achievement (Δ = 0.42) may be the most significant value when viewed from the new curriculum developers perspective. The fact that achievement scores do not plummet when the emphasis on content is reduced, but in fact, actually increase, could be considered a significant accomplishment if substantial gains in other performance areas such as student attitude and problem solving and process skills are also realized!

TABLE 14

EFFECT SIZE DATA FOR CRITERION CLUSTERS BY CURRICULUM PROFILE: EMPHASIS ON INQUIRY

CLUSTER	RATING	N	MEAN Δ	MINIMUM Δ	MAXIMUM Δ	S.D.	t-value
Achievement	High	16	0.41	-0.06	2.41	0.64	2.58*
	Low	106	0.36	-1.04	4.18	0.77	5.03*
Perception	High	18	0.13	-0.82	0.82	0.47	1.16
	Low	18	0.42	-0.81	1.75	0.51	3.46*
Process Skills	High	12	0.47	0.11	1.20	0.30	5.44*
	Low	35	0.36	-0.62	2.50	0.68	3.14*
Analytic Skills	High	2	0.05	-0.18	0.29	0.33	0.23
	Low	31	0.27	-0.36	1.44	0.46	3.26*
Related Skills	High	25	0.42	-0.09	4.48	0.88	2.42*
	Low	23	0.06	-0.50	0.75	0.29	1.02
Other Areas	High	10	0.27	-0.55	0.86	0.50	1.70
	Low	10	0.37	-0.70	1.50	0.57	2.05

*Value is significant at the *a priori* alpha level (0.05).

TABLE 15

EFFECT SIZE DATA FOR CRITERION CLUSTERS BY CURRICULUM PROFILE: EMPHASIS ON PROCESS SKILLS

CLUSTER	RATING	N	MEAN Δ	MINIMUM Δ	MAXIMUM Δ	S.D.	t-value
Achievement	High	103	0.39	-1.04	4.18	0.77	5.09*
	Low	16	0.35	-0.52	2.47	0.66	2.32*
Perceptions	High	28	0.28	-0.82	1.75	0.52	2.85*
	Low	8	0.25	-0.81	0.76	0.49	1.45
Process Skills	High	33	0.50	-0.33	2.50	0.65	4.39*
	Low	14	0.12	-0.62	0.52	0.37	1.22
Analytic Skills	High	20	0.38	-0.27	1.44	0.52	3.33*
	Low	13	0.06	-0.36	0.44	0.23	0.92
Related Skills	High	45	0.27	-0.50	4.48	0.71	2.54*
	Low	3	0.01	-0.10	0.11	0.10	0.22
Other Areas	High	17	0.35	-0.55	1.50	0.50	2.87*
	Low	3	0.16	-0.70	0.74	0.76	0.36

*Value is significant at the *a priori* alpha level (0.05).

TABLE 16

EFFECT SIZE DATA FOR CRITERION CLUSTERS BY CURRICULUM PROFILE: EMPHASIS ON LABORATORY

CLUSTER	RATING	N	MEAN Δ	MINIMUM Δ	MAXIMUM Δ	S.D.	t-value
Achievement	High	111	0.32	-1.04	2.70	0.62	5.56*
	Low	11	0.96	-0.42	4.18	1.52	2.10
Perceptions	High	31	0.24	-0.82	1.05	0.42	3.17*
	Low	5	0.47	-0.81	1.75	0.93	1.14
Process Skills	High	45	0.34	-0.62	2.50	0.53	4.33*
	Low	2	1.39	0.34	2.45	1.49	1.32
Analytic Skills	High	30	0.21	-0.36	1.41	0.42	2.79*
	Low	3	0.69	0.21	1.44	0.65	1.85
Related Skills	High	48	0.25	-0.50	4.48	0.69	2.54*
	Low	0	---	---	---	---	---
Other Areas	High	20	0.32	-0.70	1.50	0.52	2.73*
	Low	0	---	---	---	---	---

*Value is significant at the *a priori* alpha level (0.05).

TABLE 17

EFFECT SIZE DATA FOR CRITERION CLUSTERS BY CURRICULUM PROFILE: EMPHASIS ON INDIVIDUALIZATION

CLUSTER	RATING	N	MEAN Δ	MINI-MUM Δ	MAXI-MUM Δ	S.D.	t-value
Achievement	High	11	0.48	-0.06	2.41	0.76	2.07
	Low	111	0.37	-1.04	4.18	0.75	5.19*
Perceptions	High	16	0.11	-0.82	0.82	0.50	0.92
	Low	20	0.40	-0.81	1.75	0.49	3.66*
Process Skills	High	14	0.44	0.11	1.20	0.28	5.86*
	Low	33	0.36	-0.62	2.50	0.70	3.00*
Analytic Skills	High	2	0.11	-0.07	0.29	0.25	0.61
	Low	31	0.26	-0.36	1.44	0.46	3.20*
Related Skills	High	16	0.17	-0.10	0.54	0.16	4.29*
	Low	32	0.29	-0.50	4.48	0.83	1.98*
Other Areas	High	11	0.31	-0.55	0.86	0.50	2.08
	Low	9	0.33	-0.70	1.50	0.59	1.68

*Value is significant at the *a priori* alpha level (0.05).

TABLE 18

EFFECT SIZE DATA FOR CRITERION CLUSTERS BY CURRICULUM PROFILE: EMPHASIS ON CONTENT

CLUSTER	RATING	N	MEAN Δ	MINIMUM Δ	MAXIMUM Δ	S.D.	t-value
Achievement	High	107	0.37	-1.04	4.18	0.77	5.07*
	Low	15	0.42	-0.27	2.41	0.67	2.45*
Perceptions	High	18	0.42	-0.81	1.75	0.51	3.48*
	Low	18	0.12	-0.82	0.82	0.47	1.14
Process Skills	High	35	0.36	-0.62	2.50	0.68	3.12*
	Low	12	0.48	0.12	1.20	0.29	5.66*
Analytic Skills	High	32	0.27	-0.36	1.44	0.45	3.44*
	Low	1	-0.27	-0.27	-0.27	---	---
Related Skills	High	23	0.06	-0.50	0.75	0.29	1.02
	Low	25	0.42	-0.09	4.48	0.88	2.42
Other Areas	High	7	0.33	-0.70	1.50	0.68	1.28
	Low	13	0.31	-0.55	0.86	0.45	2.51*

*Value is significant at the *a priori* alpha level (0.05).

TABLE 19

CORRELATIONS BETWEEN CURRICULUM PROFILE RATINGS AND EFFECT SIZES CALCULATED FROM STUDENT PERFORMANCE DATA

PERFORMANCE(S)	N	INQUIRY	PROCESS	LABORATORY	INDIVIDUAL	CONTENT
Achievement	122	0.05	0.06	-0.07	0.07	0.04
Perceptions	36	-0.39*	-0.26*	-0.33*	-0.39*	-0.12
Process Skills	47	0.05	0.16	-0.07	0.04	-0.03
Analytic Skills	33	0.00	0.34*	-0.07	0.01	0.16
Related Skills	48	0.28*	0.09	0.00	-0.08	-0.28*
Other Areas	20	-0.11	0.03	-0.07	-0.05	-0.02
Composite	306	-0.02	0.04	-0.17*	-0.02	0.01

*Value is significant at the *a priori* alpha level (0.05).

In studies where grade level is specified, how do students exposed to new science curricula compare to students in traditional science courses?

Data are presented in Table 20 which address the question of curriculum effectiveness by grade level. Conventional grade level groupings (i.e., elementary (K-6), junior high (7-9), high school (10-12), and post secondary) are used for two reasons: (1) specific grade level data are not available in the majority of studies reported, and (2) the limited numbers of studies in many of the specific grades prohibits meaningful quantitative synthesis.

TABLE 20

EFFECT SIZE DATA FOR CRITERION CLUSTERS BY GRADE LEVEL ACROSS ALL NEW SCIENCE CURRICULA

CLUSTER	N	MEAN Δ	MINI-MUM Δ	MAXI-MUM Δ	S.D.	t-value
Elementary (K-6)						
Achievement	27	0.37	-0.57	2.47	0.74	2.64*
Perceptions	29	0.28	-0.82	0.83	0.46	3.28*
Process Skills	16	0.56	-0.02	2.50	0.59	3.84*
Analytic Skills	1	0.06	0.06	0.06	0.00	---
Related Skills	37	0.17	-0.41	1.04	0.27	3.84*
Other Areas	14	0.32	-0.55	1.50	0.55	2.22*
Composite	124	0.31	-0.82	2.50	0.52	6.55*
Junior High (7-9)						
Achievement	13	0.23	-0.33	0.86	0.34	2.46*
Perceptions	11	0.59	0.17	1.12	0.31	6.14*
Process Skills	18	0.23	-0.62	0.73	0.39	2.49*
Analytic Skills	14	0.02	-0.36	0.44	0.23	0.32
Related Skills	9	0.68	-0.10	4.48	1.46	1.41
Other Areas	7	0.33	-0.70	0.74	0.48	1.84
Composite	72	0.31	-0.70	4.48	0.62	4.24*
High School (10-12)						
Achievement	83	0.37	-1.04	4.18	0.80	4.29*
Perceptions	9	0.44	-0.81	1.75	0.70	1.90
Process Skills	19	0.43	-0.33	2.45	0.68	2.77*
Analytic Skills	19	0.42	-0.18	1.44	0.50	3.66*
Related Skills	2	-0.23	-0.50	0.04	0.38	-0.85
Composite	132	0.38	-1.04	4.18	0.73	6.06*

continued

TABLE 20 (Continued)

CLUSTER	N	MEAN Δ	MINIMUM Δ	MAXIMUM Δ	S.D.	t-value
Post-Secondary						
Achievement	5	0.47	0.13	0.82	0.27	3.91*
Perceptions	1	0.15	0.15	0.15	0.00	---
Process Skills	2	0.09	0.08	0.10	0.01	9.00
Analytic Skills	1	0.23	0.23	0.23	0.00	---
Composite	9	0.32	0.08	0.82	0.26	3.70*

*Value is significant at the *a priori* alpha level (0.05).

A rough estimate of new curriculum effectiveness by grade level is available in the "composite" line entries of Table 20. The data for all the different criterion variables are treated as one composite variable, i.e., student performance. The data show that students participating in new science curricula performed better than their traditional course counterparts by 0.31 to 0.38 standard deviations across all performance measures. Thus, the average student in new science curricula (by grade level) exceeded the performance of 62-65% of the students in traditional courses.

The detailed data in Table 20 show the effect sizes by criterion clusters. Data were available to calculate approximately 125 effect sizes for both elementary grade level and high school studies but only 72 for the junior high school level. Similarly, data were available for 11 post-secondary effect size calculations. These post-secondary effect size calculations were for study situations in which no modifications were made to the curricula being used and those students had not had a previous course in that science discipline. Thus, reasonable comparisons are able to be made.

Among the more interesting results in Table 20 are the significant differences for process skills at the elementary level, perceptions at the junior high school level, and achievement at the post-secondary level. The process skill area was targeted as a critical area among elementary school science curriculum developers. Based on the Table 20 data, that goal is being achieved. At the junior high level student attitudes were considered a prime target (i.e., getting students to like science). Here again, the new programs show their effectiveness.

What curriculum designers had not expected was the success of some programs in the post-secondary arena. The data show that when new science materials are used with junior college and beginning college students, achievement is enhanced.

When grouped by gender, how do students exposed to new science curricula compare to students in traditional science courses?

A recurring question in science education deals with the sex-bias of certain science materials and even entire subject areas. There is an intuitive notion among lay people and educators that males gravitate toward selected science areas on a random basis but, that women tend to be attracted to non-quantitative areas of science, if they are attracted toward science at all. Following the intuitive logic one step further, the reason cited for female aversion to science is their poor performance in science related areas. The question of sex-bias in the new curriculum materials is not one that this meta-analysis can answer with the type of data available. However, there are some data that deal with the question of student gender and performance in the new science curricula compared to the traditional courses. The data are presented in Table 21 and shown graphically in Figure 2.

The data in Table 21 are grouped according to the make-up of the student populations sampled in the research studies coded. If the percentage of females was reported as less than 25% of the total sample studied, the study was classified as predominantly male. If the percentage of females was reported as greater than 75%, the study was classified as predominantly female. Male/female percentages between 25% and 75% were classified as a mixed group. For the composite data of Table 21 the performance criteria were collapsed across all performance factors to provide a gross indicator of science curriculum-student gender interaction.

The breakdown of the data in Table 21 by sample type is intuitively interesting. Only 19 effect sizes were calculated for samples with more than 75% females while 123 were calculated for predominantly male samples. The composite performance results, however, show clearly that predominantly male and predominantly female samples performed equally well, about a quarter standard deviation better than

TABLE 21

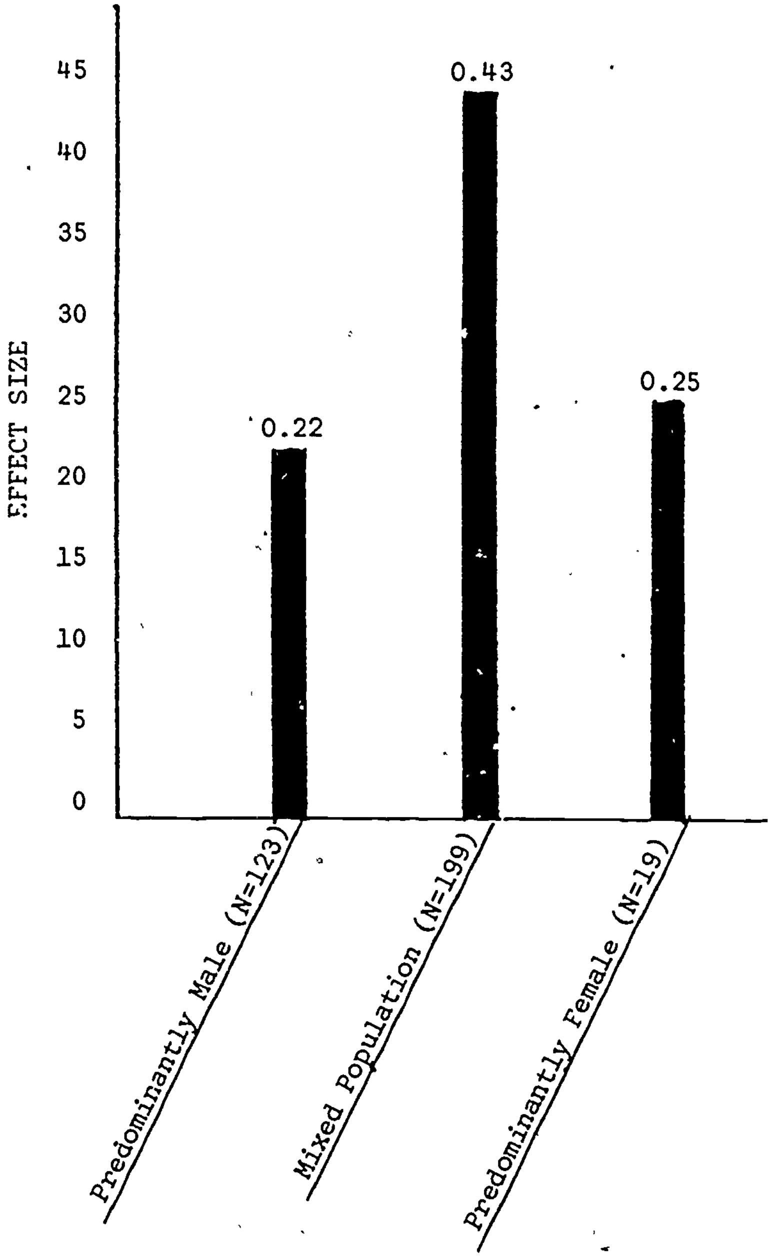
EFFECT SIZE DATA FOR CRITERION VARIABLES BY STUDENT GENDER ACROSS ALL NEW SCIENCE CURRICULA

CLUSTER	N	MEAN Δ	MINIMUM Δ	MAXIMUM Δ	S.D.	t-value
Male Sample						
Achievement	58	0.25	-1.04	2.70	0.49	3.96*
Perceptions	12	-0.02	-0.82	0.76	0.56	-0.13
Process Skills	18	0.16	-0.44	1.19	0.36	1.96
Analytic Skills	21	0.30	-0.18	1.41	0.40	3.45*
Related Skills	8	0.01	-0.22	0.28	0.14	0.33
Other Areas	6	0.47	-0.09	0.86	0.39	2.96*
Composite	123	0.22	-1.04	2.70	0.45	5.31*
Mixed Sample						
Achievement	68	0.45	-0.57	4.18	0.88	4.28*
Perceptions	34	0.51	-0.32	1.75	0.38	7.76*
Process Skills	33	0.52	-0.62	2.50	0.64	4.65*
Analytic Skills	9	0.31	-0.27	1.44	0.56	1.67
Related Skills	40	0.30	-0.50	4.48	0.74	2.55*
Other Areas	15	0.27	-0.70	1.50	0.56	1.88
Composite	199	0.43	-0.70	4.48	0.71	8.44*
Female Sample						
Achievement	4	0.55	-0.52	1.65	0.88	1.25
Perceptions	5	0.32	-0.40	0.82	0.45	1.58
Process Skills	5	0.29	-0.05	0.52	0.23	2.80*
Analytic Skills	5	-0.10	-0.36	0.21	0.24	-0.94
Composite	19	0.25	-0.52	1.65	0.50	2.15*

*Value is significant at the *a priori* alpha level (0.05).

FIGURE 2

BAR GRAPH OF THE MEAN EFFECT SIZES FOR COMPOSITE PERFORMANCE BY GENDER



STUDENT GENDER: ALL CRITERION VARIABLES ACROSS ALL NEW SCIENCE CURRICULA their traditional course comparison groups. What is not easily explained is the substantially greater mean effect size for the mixed groups. Perhaps there is a social dimension to learning and liking science that must be accounted for in the classroom.

The breakdown of the data in Table 21 reveals two interesting features. The first deals with the consistently higher effect size pattern of the mixed sample studies. On almost every criterion measure the mixed group samples produced more positive differences. The second feature deals with the analytic skills data for the female samples. The -0.10 effect size represents the largest negative result encountered in the new science curricula-traditional course comparisons.

An ANOVA was conducted on the effect size data grouped by individual performance criterion cluster and by composite performance by gender. The difference in overall student performance between the predominantly male samples and the mixed samples was statistically significant on the basis of a Duncan's Multiple Range Test, at the alpha 0.05 level. However, no significant differences between the groupings were found for the individual criterion clusters. A summary of the ANOVA data appears in Table 22.

In studies where school type is specified, how do students exposed to new science curricula compare to students in traditional science courses?

Effect size data for student performance in new science programs versus traditional courses grouped by school type are presented in

TABLE 22

ANOVA SUMMARY FOR EFFECT SIZE DATA GROUPED BY CRITERION CLUSTER AND SAMPLE GENDER

SOURCE	df	SS	MS	F-value
Model	15	7.90	0.53	1.35
Cluster	5	1.14	0.22	0.59
Gender	2	3.19	1.59	4.09*
Cluster*Gender	8	3.09	1.54	0.99
Error	325	126.91	0.39	
Corrected Total	340			

*Value is significant at the *a priori* alpha level (0.05).

Table 23. Since information regarding the school type from which samples were drawn was not available in all studies reviewed, the number of effect sizes included in the analysis is substantially reduced from previous analyses. Assuming the uncodable studies would disburse themselves equally among the three school type categories (a conservative estimate), it is clear that the bulk of research conducted on questions of curriculum effectiveness is done in suburban schools.

The composite data in Table 23 indicate that the new science curricula apparently impacted suburban and urban schools more positively than the rural schools. The breakdown of effect size data by criterion clusters magnify this composite data disparity.

TABLE 23

EFFECT SIZE DATA FOR CRITERION CLUSTERS BY SCHOOL TYPE ACROSS ALL NEW SCIENCE CURRICULA

CLUSTER	N	MEAN Δ	MINIMUM Δ	MAXIMUM Δ	S.D.	t-value
Rural						
Achievement	9	0.34	0.04	0.71	0.25	3.94*
Perceptions	9	-0.07	-0.82	0.82	0.58	-0.40
Process Skills	6	0.45	0.15	0.73	0.23	4.65*
Other Areas	1	0.06	0.06	0.06	0.00	---
Composite	25	0.20	-0.82	0.82	0.44	2.28*
Suburban						
Achievement	72	0.41	-1.04	4.18	0.85	4.11*
Perceptions	19	0.46	0.00	1.75	0.40	5.04*
Process Skills	13	0.50	-0.62	2.45	0.85	2.11*
Analytic Skills	17	0.27	-0.27	1.44	0.45	2.45*
Related Skills	34	0.30	-0.41	4.48	0.79	2.22*
Other Areas	13	0.37	-0.55	1.50	0.56	2.37*
Composite	168	0.38	-1.04	4.48	0.74	6.72*
Urban						
Achievement	4	0.81	0.20	1.08	0.41	3.95*
Perceptions	2	0.64	0.60	0.69	0.06	14.33*
Process Skills	12	0.24	-0.44	1.19	0.44	1.92
Analytic Skills	11	0.17	-0.36	1.41	0.47	1.19
Related Skills	2	0.41	0.08	0.75	0.47	1.24
Other Areas	1	0.86	0.86	0.86	0.00	---
Composite	32	0.34	-0.44	1.41	0.47	4.13*

*Value is significant at the *a priori* alpha level (0.05).

Specifically, on achievement measures the rural school mean is the lowest of the three groups though not substantially different from the suburban mean. However, on measures of student perceptions, the rural school data are more than a half standard deviation lower than both the suburban and urban groups. An ANOVA was performed on the effect size data to test the significance of the differences for each of the performance criterion groupings across the three school types. In essence a one-way ANOVA of effect size data is a test of the interaction of new curricula and school type on student performance. The differences were not statistically significant at the $\alpha \leq 0.05$ level.

When grouped by socio-economic status, how do students exposed to new science curricula compare to students in traditional science courses?

The data in Table 24 indicate that very few studies of new science curricula have been conducted in which student socio-economic status was isolated as a study variable. The extremely small number of studies completed on low socio-economic student samples (N=4) diminishes the power of quantitative synthesis techniques considerably. The distribution of the 19 effect sizes calculated for the high SES samples also limits meaningful discussion regarding the achievement studies analysis at the criterion cluster level. Nonetheless, the high SES achievement data are interesting. The achievement (Δ = 1.00) and the composite data (Δ = 0.99) effect sizes for the high SES students

TABLE 24

EFFECT SIZE DATA FOR CRITERION CLUSTERS BY STUDENT SOCIO-ECONOMIC STATUS ACROSS ALL NEW SCIENCE CURRICULA

CLUSTER	N	MEAN Δ	MINI-MUM Δ	MAXI-MUM Δ	S.D.	t-value
Low SES						
Achievement	1	1.08	---	---	0.00	---
Process Skills	1	0.64	---	---	0.00	---
Related Skills	2	0.41	0.08	1.08	0.41	1.24
Composite	4	0.63	0.08	1.08	0.41	3.06*
Mid-SES						
Achievement	105	0.27	-0.57	2.47	0.49	5.68*
Perceptions	49	0.32	-0.82	1.12	0.44	5.16*
Process Skills	49	0.33	-0.62	2.50	0.46	5.00*
Analytic Skills	33	0.23	-0.36	1.44	0.42	3.10*
Related Skills	46	0.24	-0.50	4.48	0.70	2.38*
Other Areas	20	0.31	-0.70	1.50	0.52	2.70*
Composite	302	0.28	-0.82	4.48	0.51	9.71*
High SES						
Achievement	11	1.00	-0.26	4.18	1.59	2.10*
Perceptions	2	1.40	1.05	1.75	0.49	4.00
Process Skills	4	1.00	-0.33	2.45	1.31	1.52
Analytic Skills	2	0.50	-0.08	1.08	0.82	0.86
Composite	19	0.99	-0.33	4.18	1.34	3.21*

*Value is significant at the *a priori* alpha level (0.05).

are among the highest mean values for any analysis by group in this study. An ANOVA performed on the mean effect sizes for the composite performance data shows that the high SES composite effect size mean is significantly different from the mid-SES group, on the basis of a Duncan Multiple Range Test, at the alpha 0.05 level. The ANOVA summary is presented in Table 25.

TABLE 25

ANOVA SUMMARY FOR EFFECT SIZE DATA GROUPED BY CRITERION CLUSTER AND SOCIO-ECONOMIC STATUS

SOURCE	df	SS	MS	F-value
Model	12	10.72	0.89	2.51*
Cluster	5	0.57	0.11	0.32
SES	2	5.31	2.66	7.48*
Cluster*SES	5	0.89	0.18	0.50
Error	312			
Corrected Total	324			

*Value is significant at the *a priori* alpha level (0.05).

When teachers receive inservice training with particular new science curriculum, how do students exposed to the science program compare to students in traditional courses?

The inservice data in Table 26 are extremely interesting in light of the debates and discussions during the past 25 years regarding the cost-effectiveness of in-service teacher education. Unfortunately, the data for this analysis are incomplete, as detailed information regarding inservicing was available only in about 30% of the studies coded. These studies yielded 126 effect sizes. A summary of effect size data by criterion cluster for studies not reporting the inservice backgrounds of participating teachers is also reported in Table 26 to facilitate a more meaningful discussion of the known data.

An examination of the composite and criterion cluster effect size data reveals a striking difference in overall student performance. On every measure where data are available for the inservice versus no inservice summaries the effect sizes are higher for the no inservice studies. Even if one were to assume that the effects of inservice education would only result in improved achievement scores for students taking courses from such teachers, the no inservice effect size for achievement ($\Delta = 0.46$) is considerably higher than the achievement effect size ($\Delta = 0.22$) for students taking courses from teachers who did not receive such education.

When teachers receive special instruction in the use of materials for a particular science curriculum or method prior to receiving teacher certification (preservice instruction), how do students exposed to those science programs compare to students in traditional courses?

When considering the inservice data of Table 26 it is difficult to know exactly how many, or what percentage, of the teachers may have received inservice instruction in studies where that information was not reported. Based upon the years of experience data coded however, approximately 65% of the teachers participating in the

TABLE 26

EFFECT SIZE DATA FOR CRITERION CLUSTERS BY INSERVICE EXPERIENCE ACROSS ALL NEW SCIENCE CURRICULA

CLUSTER	N	MEAN Δ	MINIMUM Δ	MAXIMUM Δ	S.D.	t-value
Inservice						
Achievement	40	0.22	-0.52	1.65	0.41	3.50*
Perceptions	19	0.16	-0.82	0.82	0.48	1.47
Process Skills	27	0.32	-0.62	2.50	0.55	3.02*
Analytic Skills	15	0.07	-0.36	0.44	0.22	1.36
Related Skills	5	0.12	-0.22	0.57	0.30	0.91
Other Areas	6	0.57	-0.09	1.50	0.57	2.47*
Composite	112	0.23	-0.82	2.50	0.45	5.44*
No Inservice						
Achievement	9	0.46	-0.13	0.92	0.39	3.53*
Perceptions	2	0.64	0.60	0.69	0.06	14.33*
Process Skills	1	0.32	---	---	0.00	---
Analytic Skills	2	0.62	0.49	0.75	0.18	4.77
Composite	14	0.50	-0.13	0.92	0.32	5.72*
Data Not Reported						
Achievement	81	0.43	-1.04	4.18	0.87	4.46*
Perceptions	30	0.48	-0.81	1.75	0.47	5.64*
Process Skills	28	0.45	-0.33	2.45	0.58	4.13*
Analytic Skills	18	0.34	-0.27	1.44	0.54	2.68*
Related Skills	43	0.26	-0.50	4.48	0.72	2.43*
Other Areas	15	0.23	-0.70	0.86	0.47	1.88
Composite	215	0.38	-1.04	4.48	0.71	8.01*

*Value is significant at the *a priori* alpha level (0.05).

studies "not reporting' preservice background graduated prior to 1960. This rules out the possibility of any preservice training for such teachers. With this in mind, roughly 15% of the studies reported teachers receiving some form of preservice instruction.

The data in Table 27, like the inservice data, form a pattern showing a less positive impact on student performance when teachers involved in the studies received preservice instruction. Possible reasons for the preservice and inservice performance data are discussed in the summary statements of this report.

Study Characteristics

When the level of internal validity is accounted for in studies meta-analyzed, how do students exposed to new science curricula compare to students in traditional science courses?

Critics of meta-analysis express concern regarding the problem of combining results from both "good" and "poor" studies. Certainly this is a valid criticism if the collective results of studies rated "good" are significantly different from those rated "poor." But, as Glass (1980) points out, ". . . if 'good' and 'poor' studies do not differ greatly in their findings, a large data base (all studies regardless of quality) is much to be preferred over a small data base (only the "good" studies). The larger data base can be more readily subdivided to answer specific sub-questions that are inevitably provoked by the answers to the general questions . . ." (p. 286).

TABLE 27

EFFECT SIZE DATA FOR CRITERION CLUSTERS BY PRESERVICE EXPERIENCE ACROSS ALL NEW SCIENCE CURRICULA

CLUSTER	N	MEAN Δ	MINI-MUM Δ	MAXI-MUM Δ	S.D.	t-value
Preservice						
Achievement	18	0.20	-0.49	1.09	0.36	2.35*
Perceptions	4	0.39	0.21	0.66	0.20	3.79*
Process Skills	9	0.22	-0.62	0.73	0.41	1.61
Composite	31	0.23	-0.62	1.09	0.36	3.60*
No Preservice						
Achievement	9	0.30	-0.13	0.92	0.39	2.25*
Perceptions	12	0.09	-0.82	0.82	0.59	0.55
Process Skills	2	1.41	0.32	2.50	1.54	1.29
Analytic Skills	5	0.27	0.01	0.75	0.32	1.90
Related Skills	1	0.57	---	---	0.00	---
Other Areas	1	1.50	---	---	0.00	---
Composite	30	0.33	-0.82	2.50	0.65	2.82*
Data Not Reported						
Achievement	103	0.40	-1.04	4.18	0.80	5.15*
Perceptions	35	0.46	-0.81	1.75	0.44	6.14*
Process Skills	45	0.37	-0.44	2.45	0.50	4.98*
Analytic Skills	30	0.24	-0.36	1.44	0.46	2.85*
Related Skills	47	0.24	-0.50	4.48	0.69	2.43*
Other Areas	20	0.27	-0.70	0.86	0.45	2.68*
Composite	280	0.35	-1.04	4.48	0.65	9.13*

*Value is significant at the *a priori* alpha level (0.05).

Studies included in this meta-analysis were rated on several design features. One of these features was the overall internal validity of the study which the coders rated as: (1) Low (intact and highly dissimilar groups), (2) Medium (random samples or matched samples with some threats to internal validity), or (3) High (random samples with low mortality). Summary effect size data for the three rated levels of internal validity are presented in Table 28.

An examination of the composite effect size means in Table 28 reveals a range of 0.05 standard deviations between studies at the extremes of judged internal validity. A one-way ANOVA of the effect size data by rated internal validity, revealed no significant differences in the composite effect size means of the three validity rankings. A further analysis of effect size means for each criterion cluster by judged internal validity also revealed no significant differences. It is safe to assume, therefore, that any conclusions based on sub-groupings of study results reported herein are not weakened by level of internal validity of the original research studies analyzed.

When the type of criterion measure used is considered in the meta-analysis, how do students exposed to new science curricula compare to students in traditional courses?

A particular threat to a study's internal validity concerns the instrumentation used in measuring the dependent variables, in this case, student performance. Unvalidated, experimenter-made tests pose a threat to validity because of the high risk of experimenter bias in

TABLE 28

EFFECT SIZE DATA FOR CRITERION CLUSTERS BY RATED LEVEL OF INTERNAL VALIDITY

CLUSTER	N	MEAN Δ	MINIMUM Δ	MAXIMUM Δ	S.D.	t-value
Low Internal Validity						
Achievement	52	0.33	-0.52	2.70	0.67	3.59*
Perceptions	17	0.51	-0.32	1.12	0.35	6.01*
Process Skills	10	0.58	-0.10	2.50	0.71	2.59*
Analytic Skills	7	0.28	0.01	1.22	0.42	1.76
Related Skills	16	0.21	-0.22	0.75	0.25	3.39*
Other Areas	8	0.16	-0.55	0.81	0.53	0.07
Composite	110	0.35	-0.55	2.70	0.56	6.49*
Medium Internal Validity						
Achievement	66	0.40	-1.04	4.18	0.84	3.91*
Perceptions	30	0.27	-0.82	1.75	0.56	2.66*
Process Skills	40	0.33	-0.62	2.45	0.55	3.83*
Analytic Skills	27	0.25	-0.36	1.44	0.46	2.82*
Related Skills	32	0.27	-0.50	4.48	0.83	1.86
Other Areas	10	0.37	-0.70	0.86	0.42	2.79*
Composite	205	0.33	-1.04	4.48	0.68	6.87*
High Internal Validity						
Achievement	12	0.35	-0.02	0.86	0.28	4.29*
Perceptions	4	0.47	0.37	0.55	0.07	12.37*
Process Skills	5	0.35	0.26	0.63	0.15	5.03*
Analytic Skills	1	-0.07	---	---	0.00	---
Other Areas	3	0.61	0.06	1.50	0.77	1.38
Composite	25	0.38	-0.07	1.50	0.33	5.86*

*Value is significant at the *a priori* alpha level (0.05).

the construction or selection of test items. While there is no test which is totally unbiased, a conservative approach to the resolution of the test bias question in a research synthesis study is to segregate those studies using standardized tests for closer scrutiny. The results of the meta-analysis on student performance data by standardized test versus other forms is presented in Table 29.

The data in Table 29 indicate that student performance results do not appear to be influenced by the type of test used. No significant differences were revealed in the comparison of composite performance data nor on criterion cluster data between standardized and other test forms.

When length of treatment is isolated as a factor, how do students exposed to new science curricula compare to students in traditional science courses?

Campbell and Stanley (1966) define internal validity as -- "the basic minimum without which any experiment is uninterpretable" (p. 5). One of the major threats to a study's internal validity deals with treatment fidelity; in other words, are the treatment conditions which characterize the comparison groups discernable and reasonable. Few studies of new science curricula reported information from which treatment fidelity could be judged. However, considering the nature of the treatment condition of interest in this meta-analysis (i.e., student exposure to new science curricula versus traditional programs), the length of exposure to the programs constitutes a reasonable approximation to the question of treatment fidelity.

TABLE 29

EFFECT SIZE DATA FOR CRITERION CLUSTERS BY TYPE OF TEST USED

CRITERION	N	MEAN Δ	MINIMUM Δ	MAXIMUM Δ	S.D.	t-value
Standardized						
Achievement	73	0.35	-1.04	4.18	0.76	3.94*
Perceptions	34	0.34	-0.82	1.75	0.49	3.97*
Process Skills	38	0.33	-0.62	2.45	0.53	3.87*
Analytic Skills	35	0.24	-0.36	1.44	0.44	3.29*
Related Skills	31	0.23	-0.41	4.48	0.81	1.58
Other Areas	7	0.59	-0.09	1.50	0.50	3.12*
Composite	218	0.32	-1.04	4.48	0.64	7.35*
Other Forms						
Achievement	57	0.39	-0.52	2.70	0.69	4.26*
Perceptions	17	0.43	-0.81	1.12	0.48	3.71*
Process Skills	18	0.50	-0.17	2.50	0.62	3.44*
Related Skills	17	0.29	-0.50	1.04	0.38	3.17*
Other Areas	14	0.19	-0.70	0.86	0.48	1.52
Composite	123	0.37	-0.81	2.70	0.59	7.01*

*Value is significant at the *a priori* alpha level (0.05).

Table 30 contains summary effect size data for studies grouped by the length of study. Four levels of treatment duration were chosen for analysis: (1) less than 10 weeks, (2) between 10 and 20 weeks, (3) between 21 and 36 weeks, and (4) longer than 36 weeks.

TABLE 30

EFFECT SIZE DATA FOR CRITERION CLUSTERS BY LENGTH OF TREATMENT

CRITERION	N	MEAN Δ	MINIMUM Δ	MAXIMUM Δ	S.D.	t-value
<u>Less than 10 Weeks</u>						
Achievement	6	0.55	-0.49	2.47	1.01	1.34
Perceptions	2	0.43	0.35	0.51	0.11	5.38
Analytic Skills	1	0.01	---	---	0.00	---
Related Skills	1	0.34	---	---	0.00	---
Other Areas	6	0.05	-0.55	0.81	0.53	0.27
Composite	16	0.30	-0.55	2.47	0.70	1.73
<u>Between 10 and 20 Weeks</u>						
Achievement	14	0.22	-0.03	0.92	0.29	2.86*
Perceptions	19	0.21	-0.82	1.12	0.58	1.59
Process Skills	11	0.37	0.08	0.70	0.17	6.98*
Analytic Skills	3	0.43	0.06	0.75	0.34	2.15
Related Skills	12	0.49	-0.10	4.48	1.27	1.34
Other Areas	2	0.89	0.29	1.50	0.85	1.48
Composite	61	0.33	-0.82	4.48	0.68	3.82*
<u>Between 21 and 36 Weeks</u>						
Achievement	94	0.39	-1.04	4.18	0.79	4.80*
Perceptions	23	0.46	-0.81	1.75	0.47	4.75*
Process Skills	43	0.40	-0.62	2.50	0.63	4.12*
Analytic Skills	30	0.23	-0.36	1.44	0.46	2.76*
Related Skills	21	0.24	-0.50	1.04	0.33	3.32*
Other Areas	6	0.19	-0.70	0.49	0.45	1.07
Composite	217	0.36	-1.04	4.18	0.65	8.16*

continued

TABLE 30 (Continued)

CRITERION	N	MEAN Δ	MINIMUM Δ	MAXIMUM Δ	S.D.	t-value
Longer than 36 Weeks						
Achievement	14	0.32	-0.57	1.65	0.57	2.11*
Perceptions	6	0.50	0.21	0.76	0.19	6.29*
Process Skills	1	0.32	---	---	0.00	---
Related Skills	12	0.00	-0.41	0.33	0.25	0.02
Other Areas	5	0.40	-0.09	0.74	0.35	2.55
Composite	38	0.26	-0.57	1.65	0.43	3.69*

*Value is significant at the *a priori* alpha level (0.05).

The data in Table 30 show that length of treatment appears to have no effect upon the composite performance data. The values range from Δ = 0.26 for studies longer than 36 weeks to Δ = 0.36 for studies between 21 and 36 weeks long. An ANOVA of both the composite performance data and performance cluster data revealed no significant differences in mean effect size values. It is interesting to note, however, that composite performance effect size data show a pattern of positive increases for studies spanning periods less than 10 weeks up to 36 weeks and a decline for studies conducted across two school years (greater than 36 weeks). This regression effect of composite student performance after 36 weeks' exposure contrasted to the stabilization, even strengthening of the perceptions data for treatments beyond 36 weeks (Δ = 0.50) suggests that new science curricula may have been most effective in changing student attitudes.

When studies are grouped by the form of publication, how do students exposed to new science curricula compare to students in traditional science courses?

Whether a research report is published, or where it is published, does not constitute a threat to the validity of a study nor should it be considered a source of bias. But, there is a prevailing notion among some researchers and practitioners that only significant results are publishable. In coding the results and characteristics of studies included in this meta-analysis of new science curricula effects, the primary source of the study was coded. In the event a study was published in two forms (e.g., a dissertation and a journal article), the original source of the data was recorded (in this example, the dissertation). Summary effect size data for student performance variables grouped by source, or form of publication, are presented in Table 31.

The data in Table 31 reveal no major differences in composite performance effect size when grouped according to publication form. In fact, the pattern of more pronounced effect size values for the unpublished materials contradicts the "only significant results are published" argument. It is interesting to note though, that the lion's share of research on this topic has been completed by graduate student researchers completing doctoral dissertations.

TABLE 31

EFFECT SIZE DATA FOR CRITERION CLUSTERS BY FORM OF PUBLICATION

CRITERION	N	MEAN Δ	MINIMUM Δ	MAXIMUM Δ	S.D.	t-value
Journal Articles						
Achievement	26	0.36	-1.04	2.70	0.81	2.27*
Perceptions	10	0.44	0.05	0.85	0.25	5.49*
Process Skills	10	0.39	0.08	0.70	0.23	5.23*
Analytic Skills	2	0.47	-0.27	1.22	1.05	0.64
Related Skills	11	-0.08	-0.41	0.28	0.19	-1.45
Other Areas	4	0.41	-0.09	0.86	0.44	1.86
Composite	63	0.30	-1.04	2.70	0.59	4.11*
Dissertations/Theses						
Achievement	97	0.39	-0.52	4.18	0.73	5.24*
Perceptions	34	0.36	-0.82	1.75	0.53	3.95*
Process Skills	40	0.42	-0.62	2.50	0.64	4.15*
Analytic Skills	32	0.23	-0.36	1.44	0.42	3.18*
Related Skills	25	0.20	-0.50	0.75	0.23	4.40*
Other Areas	15	0.28	-0.70	1.50	0.58	1.94
Composite	243	0.34	-0.82	4.18	0.61	8.84*
Unpublished Material						
Achievement	7	0.14	-0.46	0.54	0.43	0.86
Perceptions	7	0.29	-0.81	0.76	0.54	1.43
Process Skills	6	0.14	-0.17	0.30	0.20	1.79
Analytic Skills	1	0.15	---	---	0.00	---
Related Skills	12	0.66	-0.01	4.48	1.25	1.84
Other Areas	2	0.53	0.25	0.81	0.39	1.89
Composite	35	0.37	-0.81	4.48	0.81	2.74*

*Value is significant at the *a priori* alpha level (0.05).

SUMMARY

Literally dozens of interesting questions come to mind when the issue of curriculum effectiveness is raised. Numerous factors enter into the interpretation of data regarding even the most straightforward question. In the case of the new science curricula developed in the post-Sputnik years of the sixties and seventies, numerous studies were completed in which student performance in new science programs was compared to student performance in traditional courses. The results of any one study regarding the impact of a particular program are a matter of record. The collective results of the multiple studies conducted on several of the curricula are not so numerous. Moreover, these reports tend to be qualitative summaries which lack credibility and engender little or no confidence in the field.

The criticisms of qualitative research integration techniques are well-known. Jackson (1978) summarizes these criticisms as follows:

(1) Reviewers often ignore previous reviews on the same or similar topic.

(2) Reviewers often run the risk of sampling errors by selecting non-representative subsets of existing literature.

(3) Reviewers often use inappropriate representations of study results (e.g., whether or not results were statistically significant).

(4) Reviewers often fail to recognize and account for study characteristics which might affect results

(e.g., study sample, treatment fidelity, testing procedures).

(5) Reviewers often report so little about their review procedures it is difficult to judge the validity of the conclusions.

In an effort to tease out of the literature a concentrated mass of summative data regarding the comparison of student performance in new and traditional curricula and to avoid the pitfalls of research integration, a quantitative analysis of experimental and quasi-experimental results from the retrievable literature was performed. The quantitative integration technique used is referred to as meta-analysis (Glass, 1976). Conclusions in this report are based on data from 105 studies deemed suitable for quantitative integration from the pool of 302 studies identified. The 105 studies yielded information on 27 different new science curricula and 18 different student performance measures. Approximately 70 study characteristics were coded in reviewing each research report. Using the distribution of science curricula and student performance criteria and the collection of study characteristics, 15 major sub-questions were analyzed. A summary of results is presented below:

I. The average student exposed to new science curricula exceeded the performance of 65% of the students in traditional science courses on the aggregate criterion variable.

II. The effects of new science curricula on student performance were most impressive for the following performance criteria: creativity, laboratory techniques, attitude toward specific subject and science, and general achievement. The only negative effect size calculated for students exposed to new science curricula was for student self-concept ($\Delta = -0.08$).

III. Student overall performance scores were found to be significantly more positive for mixed student samples than with either female or male groups among the new curricula studied.

IV. Student overall performance scores were found to be significantly more positive for both high and low socio-economic students samples than for the mid-range socio-economic groups.

V. Student overall performance scores were found to be significantly more positive for student samples attending either urban or suburban schools than for rural school students among the new curricula studied.

VI. New science curricula appear to have been most effective in enhancing student process skill development at the elementary school level.

VII. New science curricula appear to have been most effective in changing student attitudes at the junior high school level.

VIII. New science curricula appear to have been most effective in enhancing student analytic skills at the high school level.

IX. New science curricula appear to have been most effective in enhancing student achievement at the post-secondary, secondary and elementary grade levels.

X. New science curricula emphasizing inquiry, process, laboratory and individualization were observed to adversely affect student perceptions about their experiences in the programs.

XI. New science curricula emphasizing process skill development were observed to adversely affect student analytic thinking skills.

XII. New science curricula emphasizing laboratory activity were observed to adversely affect overall student performance.

XIII. Of the major science curricula studied, BSCS-Blue, BSCS-Yellow, and PSSC exhibited the best overall student performance record.

XIV. Of the new curricula studied, PSSC and BSCS-Blue were found to be most effective in enhancing student achievement.

XV. Of the new curricula studied, BSCS-Blue, BSCS-Green, HSP, SCIL, and IS were found to be most effective in enhancing student perceptions.

XVI. Of the new curricula studied, BSCS-Blue, SCIS, ESS, and S-APA were found to be most effective in enhancing student process skills.

XVII. Of the new curricula studied, BSCS-Blue, PSSC and CHEM Study were found to be most effective in enhancing student analytic thinking skills.

XVIII. Of the new curricula studied, USMES was found to be most effective in enhancing student skills in related areas.

XIX. Student overall performance scores were observed to be considerably lower when teachers reported having received either inservice or preservice training in the use of the program.

The 19 summary statements must be examined carefully not only in light of the data presented here and the detail of the primary research on which the analyses were based, but also in light of what was <u>not</u> reported here or in the primary research. Perhaps the greatest uncertainty lies in the original treatment conditions themselves: was the PSSC, CHEM Study, . . . , SCIS program really being implemented according to the philosophy of the curriculum or were the materials just being used? In some studies, could the traditional treatment actually have been more of a "new curriculum" because of some innovative teacher methodologies than the new curriculum to which it was being compared? This broad question of treatment fidelity and verification is a critical issue which defies resolution but certainly can confound and distort conclusions.

For example, summary Statements X, XI, and XII dealing with curriculum emphasis on inquiry, laboratory, process, and individualization and their relationship to overall student performance are prime candidates for the distorted data file. Recall the ratings on these parameters were made by a panel of science educators. Would the teachers using the various curricula rate them similarly? Even if they rated them similarly, would they implement the curricula with these same emphases? Surely we all have seen or heard of a teacher lecturing about the inquiry method!

Similarly, the results of the inservice and preservice analyses (summary statement XIX) require careful examination. What proportion of the typical NSF inservice program was spent on learning about the new curriculum and what proportion was spent in organic chemistry, parisitology, or quantum mechanics? Should inservice and preservice programs be abandoned or written off based on these data, or, should there be a resolve to revamp the programs and change the emphasis?

BIBLIOGRAPHY

BIBLIOGRAPHY

Barr, A. J., Goodnight, J. H., Sall, J. P. and Helwig, J. T. A user's guide to SAS-76. Raleigh, North Carolina: Sparks Press, 1976.

Brown, B. B. and Webb, J. N. Valid and reliable observations of classroom behavior. Classroom Interaction Newsletter, 1968, 4, 35-38.

Butts, D. P., Bybee, R. W., Gallagher, J. J. and Yager, R. E. Assessing the current status of science education. In R. E. Yager, Crisis in science education. Science Education Center Technical Report #21. Iowa City, Iowa: The University of Iowa, 1980.

Collette, A. T. Science teaching in the secondary school: A guide for modernizing instruction. Boston: Allyn and Bacon, Inc., 1973.

Conant, J. B. The comprehensive high school: A second report to interested citizens. New York: McGraw-Hill, 1967.

Gallagher, J. J. A summary of research in science education for the years 1968-1969: Elementary school level. Journal of Research in Science Teaching, 1972, 9, 19-46.

Glass, G. V. Primary, secondary and meta-analysis of research. Educational Researcher, 1976, 5, 3-8.

Haney, R. E. The changing curriculum: Science. Washington, D.C.: National Education Association, 1966.

Haney, R. *et al.* A summary of research in science education for the years 1965-1967: Elementary school level. Research review series - science, paper 2. Columbus: ERIC Information Analysis Center for Science and Mathematics Education, 1969. (ERIC Document Reproduction Service No. ED 038 554)

Harms, N. C. and Yager, R. E. What research says to the science teacher (Volume 3). Washington, D.C.: National Science Teachers Association, 1981.

Helgeson, S. L., Blosser, P. E. and Howe, R. W. Science education (Volume 1). The status of pre-college science, mathematics, and social science education: 1955-1975. Washington, D.C.: U.S. Government Printing Office, 1978.

Helwig, J. T. and Council, K. A. (Eds.) SAS user's guide: 1979 edition. Cary, North Carolina: SAS Institute, Inc., 1979.

Hurd, P. D. Biological education in American secondary schools 1890-1960. Washington, D.C.: American Institute of Biological Sciences, 1961.

Jackson, G. B. Methods for integrative reviews. Review of Educational Research, 1980, 50 (3), 438-460.

Jacobson, W. J. Approaches to science education research: Analysis and criticism. Journal of Research in Science Teaching, 1970, 7, 217-225.

Kahl, S. and Harms, N. Project synthesis: Purpose, organization, and procedures. In N. C. Harms and R. E. Yager (Eds.), What research says to the science teacher (Volume 3). Washington, D.C.: National Science Teachers Association, 1981.

Klopfer, L. Evaluation of learning in science. In B. S. Bloom, J. T. Hastings and G. E. Madans (Eds.), Handbook of formative and summative evaluation. New York: McGraw-Hill, 1971.

Lacey, A. L. Guide to science teaching in secondary schools. Belmont, California: Wadsworth Publishing Company, 1966.

National Assessment of Educational Progress. Science: Second assessment (1972-73): Changes in science performance, 1969-73, with exercize volume and appendix (April 1977); 04-S-21, Science technical report: Summary volume (May 1977). Science: Third assessment (1976-1977): 08-S-04, Three national assessments of science: Changes in achievement, 1969-77 (June 1978): 08-S-08, The third assessment of science, 1976-77. Released exercize set (May 1978). Also some unpublished data from the 1976-77 Science Assessment, Denver, Colorado: 1870 Lincoln Street.

National Science Foundation. What are the needs in precollege science, mathematics, and social science education? Views from the field. Washington, D.C.: National Science Foundation, 1980.

Novak, J. D. A case study of curriculum change. School Science and Mathematics, 1969, 69, 375-384.

Ramsey, G. A. and Howe, R. W. An analysis of research on instructional procedures in secondary school science. Part I: Outcomes of instruction. The Science Teacher, 1969(a), 36 (3), 62-76.

Ramsey, G. A. and Howe, R. W. An analysis of research on instructional procedures in secondary school science. Part II: Instructional procedures. The Science Teacher, 1969(b), 36 (4), 72-81.

Richardson, J. S. Science teaching in secondary schools. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1964.

Rutherford, F. J. Preface. In NSF Report SE 80-9, What are the needs in precollege science, mathematics, and social science education? Views from the field. Washington, D.C.: National Science Foundation, 1980.

Schlessinger, F. R. and Helgeson, S. L. National programs in science and mathematics education. School Science and Mathematics, 1969, 69, 633-643.

Schwab, J. J. Biology teachers' handbook. New York: Wiley, 1963.

Science Policy Research Division. The National Science Foundation and pre-college science education: 1950-1975. A report prepared for the subcommittee on science, research, and technology. Washington, D.C.: U.S. Government Printing Office, 1975.

Stake, R. E. and Easley, J. Jr. Case studies in science education. Washington, D.C.: U.S. Government Printing Office, 1978.

Thurber, W. A. and Collette, A. T. Teaching science in today's secondary schools (3rd ed.). Boston: Allyn and Bacon, Inc., 1968.

Washton, N. S. Teaching science creatively in the secondary schools. Philadelphia: W. B. Saunders Company, 1967.

Weiss, I. R. Report of the 1977 national survey of science, mathematics, and social studies education. Washington, D.C.: U.S. Government Printing Office, 1978.

Welch, W. W. The impact of national curriculum projects -- The need for accurate assessment. School Science and Mathematics, 1968, 68, 225-234.

Welch, W. W. Curriculum evaluation. Review of Educational Research, 1969, 39, 429-443.

Yager, R. E. Crisis in science education. Science Education Center Technical Report #21. Iowa City, Iowa: The University of Iowa, 1980a.

Yager, R. E. Status study of graduate science education in the United States, 1960-80. Final report for NSF Grant Number 79-SP-0698, The University of Iowa, Iowa City, Iowa, 1980b.

Yager, R. E. Prologue. In N. C. Harms and R. E. Yager (Eds.), What research says to the science teacher (Volume 3). Washington, D.C.: National Science Teachers Association, 1981a.

Yager, R. E. Analysis of current accomplishments and needs in science education. Unpublished report, Science Education Center, The University of Iowa, 1981b.

Yager, R. E. The current situation in science education. (in press)

INSTRUCTIONAL SYSTEMS IN SCIENCE EDUCATION

John B. Willett
June J. M. Yamashita*

School of Education
Stanford University

---

*The order of the co-authors has been determined alphabetically.

ON BECOMING META-ANALYTICALLY LITERATE
(THOUGH PENNILESS)

As reported designs proved distressing
Reflections in Glass were a blessing
When coding a t
MS, r or p
Or rampant covariance guessing.

Though, as savants of the random statistic,
We've furthered the "cause analitique,"
To earn bread and butter
'Tis better. (than meta-)
To be a jongleur or auto-mechanique.

TABLE OF CONTENTS

LIST OF TABLES

SECTION I: SETTING UP THE META-ANALYSIS

SETTING UP THE META-ANALYSIS

Introduction and Definition of Terms

The Stanford group was assigned the question: "What are the effects of different instructional systems used in science teaching?" It was necessary initially to clarify the meaning of "systems" in order to provide as complete an analysis as possible while avoiding any overlap with the analysis performed by other study centers. The following definition was provided by the steering committee:

> An instructional system is a general plan for conducting a course over an extended period of time. It is general in that it often encompasses many aspects of a course (e.g., presentation of content, testing, size of study groups). Examples of instructional systems are: mastery learning, competency-based instruction, programmed instruction, modular instruction, mini-courses, ability grouping, team teaching, departmentalized vs. self-contained, diagnostic-prescriptive instruction, independent study/projects, computer-managed or computer assisted instruction, audio-tutorial.

An earlier draft had stated that instructional systems are "usually evaluated in an actual classroom as opposed to being evaluated in a laboratory," and "typically involve the comparison of a new learning approach with traditional instruction."

On the basis of such definitions an initial list of systems was provided at the training session in October 1980, and included those listed above. Subsequent refinements led to the designation of certain systems on the original list as "methods" or "techniques" rather than "systems," and these were then reallocated to other study centers.

The following systems were covered by the Stanford group and are reported here:

(1) Audio-Tutorial

(2) Computer-Linked, also reported separately in three categories:

  (a) Computer Assisted Instruction (CAI)

  (b) Computer Managed Instruction (CMI)

  (c) Computer Simulated Experiments (CSE)

(3) Contracts for Learning

(4) Departmentalized Elementary School

(5) Individualized Instruction

(6) Mastery Learning

(7) Media-Based Instruction, also reported separately as:

  (a) Film Instruction

  (b) Television Instruction

(8) Personalized System of Instruction (Keller PSI)

(9) Programmed Learning

  (a) Branched Programmed Learning

  (b) Linear Programmed Learning

(10) Self-Directed Study

(11) Use of Original Source Papers in the Teaching of Science

(12) Team Teaching

Each of the systems included in this report will be briefly discussed.

(1) Audio-Tutorial System. Good (1973:50) defined the Audio-Tutorial System as "a self-pacing multimedia system of instruction that features tape recorded lessons with kits of learning materials and instruction sheets for individual learning in study carrels." Descriptions given in studies evaluated here which purported to investigate this system were consistent with the above definition. Frequently, the method was referred to on the

college level as "Postlethwaite's Audio-Tutorial System."

Dr. S. N. Postlethwaite first used this system in a freshman botany class in 1961 at Purdue University, and described it as: "Audio programming of learning experiences . . . includes lectures, reading of text or other appropriate material, making observations on demonstration set-ups, doing experiments, watching movies and/or any other appropriate activities helpful in understanding the subject matter" (Postlethwaite, Novak, & Murray, 1964:6).

Audio-tutorial lessons may incorporate behavioral objectives, learning for mastery, self-pacing, and multi-media activities. Audiotapes are used "to pace students through integrated laboratory, lecture, discussion, and demonstration activities" (Nordland et al., 1972:673). (However, many studies which were coded here failed to report the exact constitution of their program, preferring simply to assume that such details were implicit in the label "Audio-tutorial instruction.")

As in other forms of individually-paced instruction, audio-tutorial systems purport to use learning time more efficiently and effectively:

> The crucial variable in [comparing A-T and traditional instruction] is not whether students under one instructional approach acquire more knowledge than under the other instructional approach but rather the analysis of learning time required to reach a given level of attainment and the quality of subsuming concepts acquired in the process [Novak 1970:782].

However, in the studies coded here, this question was largely ignored and consequently no summary on use of time is included in this report.

(2) <u>Computer-Linked Systems</u>. This category was created during the data analysis by consolidating effect sizes obtained under the next three headings.

(a) <u>Computer Assisted Instruction</u> (CAI) pertains to the use of the computer as a teaching machine. Good (1973:589) defined teaching machines as devices which ". . . control the material to which the student has access at any moment, preventing him from looking ahead or reviewing old items; . . . contain a response mechanism, that is, . . . a keyboard or selection buttons; some provision is made for knowledge of results . . . ; . . . score responses and tabulate errors." These tutoring programs sometimes (but not always) provide for student choice in content, sequencing, or type of instruction. The claimed superiority of CAI to conventional teaching derives from its supposed potential for providing immediate feedback to the student on each response and offering appropriate remediation in a manner that is often not feasible in the traditional classroom.

That not very many studies were found using computers with pre-college classes is not surprising--a recent National Science Foundation study (Weiss 1978:19) reported that only 9% of science classes in grades 10-12 ever use computers or computer terminals, although 36% of high schools have them, indicating that computers are used more for mathematics classes or for administrative purposes than for science instruction.

(b) Computer Managed Instruction (CMI), on the other hand, does not provide actual instruction for the student. Instead, the computer may be used to generate tests for students based on specific objectives, making random selections from a pool of items; to keep an up-to-date record of each student's progress in meeting learning objectives; to prescribe additional learning or remediation tasks; to plan interactively an individual student's route through pre-stored curricula, and so forth.

(c) Computer Simulated Experiments (CSE) have much potential in science instruction but are found in only a few studies at the pre-college level (Hartley, 1976:69-70). These simulations allow the student to operate with a simpler system than would actually be present in the laboratory, with unimportant or extraneous factors eliminated in the computer program. In addition, simulations allow for a wider range of student explorations in areas that would, in reality, be too dangerous, too time-consuming, or too costly.

(3) Contracts for Learning are established between an individual student and the teacher, and include the content, activities, deadlines, and methods of evaluation. Contracts would generally be a component of Self-Directed Study or other forms of independent study. A 1977 study for the National Science Foundation found that 78% of science classes never use contracts (Melton 1980:126)--which may indicate only that the remaining 22% of science classes use contracts occasionally or seldom, or for specific aspects of a course.

(4) Departmentalized Elementary School refers to the teaching of elementary school science by a specialist rather than by the typically generalist teacher. The specialist would ordinarily have a greater degree of academic training in the particular aspect of the science taught.

(5) Individualized Instruction subsumes several of the other areas described in this section, and is a catch-all term for many different approaches. In many cases, the experimental intervention used by the studies included under this system was labelled "individualized" when all students studied the same learning materials in the same sequence and their learning was evaluated in the same way; the only difference was in their pacing, with students allowed to proceed "at their own rate," using

individual packets of learning material. In contrast, Ramsey and Howe (1969: 73) offered a much broader description:

> Individualized instruction attempts to provide a complete instructional program designed explicitly for each individual, taking into account his background experience, interests, and ability.

Individualized instruction may also have been coded as Audio-Tutorial, Computer Assisted Instruction, Contracts, PSI, Programmed Instruction, or Self-Directed Study as appropriate.

Marchese (1977:699), in a literature search of individualized instruction in science, found that although much research had been conducted in other fields, very little had been reported in science. He also questioned the adequacy of instructional materials prepared for studies on individualized instruction (1977:701), and Herring et al. (1974:11) suggested an interaction of learning materials with methods of instruction may exist.

(6) Mastery Learning, as presented by Bloom in 1968, defined mastery in terms of behavioral objectives, with class instruction supplemented with feedback/correction mechanisms (Block 1971:7-8). Tests on unit objectives are followed by supplementary instruction on objectives not attained, and the student is retested until a pre-selected mastery level is achieved. Because specific levels of attainment are specified, the important variable in mastery learning is the time required to reach those levels; however, in the studies coded here this variable was largely ignored and thus appropriate conclusions cannot be reported.

(7) Media-Based Instruction: Television and Film. Studies coded as either television instruction or film instruction are those in which these forms of media provide the primary instruction rather than supplements to classroom teaching. Several television instruction studies were evaluations of a series of televised lessons which were prepared or presented at the

State Department of Education level. Many of the studies coded on film instruction used the "Harvey White Films" for Physics.

Slides and audio tapes were included on the coding sheet, and resulting effect sizes are incorporated into "Media-Based Instruction," but these two categories are not reported separately since very few effect sizes were obtained for them.

(8) Personalized System of Instruction (PSI). Frequently referred to on the college level as the "Keller Plan," PSI generally consists of the following features (Carmichael, 1976:791-2): self-paced; learning materials divided into small modules, each of which must be mastered before going on to the next; students used as graders and tutors; lack of reliance on live lectures, with printed materials being the primary form of communication. PSI has been widely criticized for the absence of the "motivating factor" that can come from live lectures and contact with the instructor (Palladino, 1979:323; Emerson, 1975:228; Kuska, 1976:505). As with other systems, the results of studies of this method of instruction may often be confounded with the value of the instructional materials specifically prepared for the investigation. A detailed study guide for each student is a crucial factor (Smith, 1976:510; Novak, 1974:15), which may not have been provided in every case.

PSI is most likely to be found at the college level, and few studies have been found at lower levels.

(9) Programmed Learning. Schramm, frequently cited for his leadership in programmed instruction, summarized what he called the essential characteristics of programmed instruction of the Skinnerian type (1962:99):

a) an ordered sequence of stimulus items,
b) to each of which a student responds in some specified way,
c) his responses being reinforced by immediate knowledge of results,

d) so that he moves in small steps,
e) therefore making few errors and practicing mostly correct responses,
f) from which he moves, by a process of successively closer approximations, toward what he is supposed to learn from the program.

In studies comparing programmed and conventional instruction, Silberman (1962:19) noted (as was also observed here) that "conditions of conventional instruction are seldom described in such reports." In fact, this lack was evident in many studies in other categories as well, implying that the salient features of conventional or traditional teaching are well known.

A 1977 National Science Foundation study reported that 71% of all science classes never use programmed instruction (Melton, 1980:126). In all probability, the remaining 29% use this form of instruction for short units. In studies which explored teachers' and students' affective responses to programmed instruction (such as the Fund for the Advancement of Education's [1964] Four Case Studies in Programed Instruction), a frequent comment was that students became bored with the materials. Teachers who intended to continue use of these materials beyond the experimental period were those who tended to use them along with other classroom activities, for remediation or enrichment, or as aids to classroom instruction rather than as a replacement. Short programmed units were found most useful when incorporated into a planned sequence of classroom activities.

Although a study may be coded as a comparison between a certain system of instruction and conventional teaching, a major part of what is being tested may be the value of the treatment protocol. A doctoral candidate using a self-developed package would be testing not only the efficacy of the instructional approach but also of the materials themselves.

Studies on programmed learning were coded as "linear" or "branched," but only five effect sizes were obtained for the latter. The small number of studies using branching is probably a result of the greater difficulty in developing such programs, since branching provides for the student to be "routed through one or more remedial sequences of frames if he misses a question or skipped ahead if he evidences mastery of content in a sequence" (Good 1973:70).

(10) Self-Directed Study. This strategy usually includes the features described as "Contracts for Learning," with students being principally responsible for "directing" their own study. However, in the studies we reviewed, students were somewhat restricted in their "self-direction": they might have a choice in the order in which they studied various units, and sometimes in the methods in which they studied the units and were evaluated, but were unlikely to have carte blanche "across the board."

(11) Source Papers. This system of teaching is based principally on the use of selected original scientific papers, documents, books, etc., rather than on the use of a school textbook. A course based on the use of source papers involves students in the finding, reading and interpretation of these original documents with or without guidance from the teacher.

(12) Team Teaching is "a type of instructional organization involving teaching personnel and the students assigned to them, in which two or more teachers are given joint responsibility for all or a significant part of the instruction of the same group of students" (Good 1973:590). As utilized in most cases, teachers shared the responsibility of large group lectures while being individually responsible for their assigned small or medium-sized groups. Teachers may alternate in presenting the large group lectures, or one teacher may be judged by the team as a superior lecturer and thus will make all presentations while the others play supporting roles and continue to handle individually their small groups. Team planning, in which those

persons teaching the same subject in a school participate, by itself does not constitute team teaching; the "joint responsibility" mentioned above is the crucial factor. In many cases, the studies reported here as investigations of the team teaching system failed to delineate what proportion of the total classroom time was spent in large group lecture, small group discussion, tutorial and so forth. However, for a particular study to have been coded as "team teaching," it was regarded as sufficient for the investigator to have labelled it thus.

Evolving the Coding Form

Jackson (1978) recommended that, in conducting integrative reviews, previous reviews on the same or similar topics be consulted prior to sampling and coding. In the case of the current meta-analysis, this step was performed by the steering committee. Then, on the basis of this consultation, a draft coding sheet was produced in Colorado. During the training session and ensuing weeks, emphasis was placed on speed in coding studies rather than on the evaluation of the instrument itself, although some modifications were made to the coding sheets during the early stages of the coding.

It may have been more appropriate, however, to involve the research assistants in the initial review, as it would have enabled them to construct a more coherent coding sheet. For example, it would have then been possible to identify the most prominent features of different forms of individualized instruction, say, and it could have been decided that every study of a system "A" would consistently have a particular group of variables coded "yes";

or, conversely, that coding a certain feature "yes" implies certain other characteristics that are not included on the coding sheet. Since this was not the case, the coding sheets evolved in a more restricted manner. Once coding had begun, major changes on the coding sheet were difficult to make due to the lack of availability of previously coded studies which were returned to Colorado for circulation to other centers.

Previous reviews would not, however, have illustrated the lack of utility of many items on the coding sheet--that few studies, for example, describe in concise terms the school community and socio-economic status of the groups, the size of the school, or student characteristics apart from IQ or other standardized ability measures; characteristics of the participating teachers, such as age, years of teaching, educational background, or even sex; and some characteristics of the experimental procedure, such as length of each lesson, class size, or initial size of the experimental groups. A great deal of unproductive time was spent in scanning reports, looking for information on these often-omitted variables; the desire to fill in as many blanks as possible on the coding sheet was somewhat compulsive, and omission of those items would have decreased the coding time per study considerably. Since it was not possible prior to coding many studies to identify these often-omitted variables, some decisions were made midway through the coding that there would be no intensive search for information that was likely to be missing.

The final coding sheet (the fourth version) consisted of the following eleven sections, each with a number of coding variables:

(1) Identification of the Study

(2) Student Identification (Treatment group; control group)

(3) Context Characteristics (Treatment group; control group)

(4) Teacher Characteristics (Treatment group; control group)
(5) Design Characteristics (Treatment group; control group)
(6) Treatment Characteristics (Treatment group; control group)
(7) Features (Treatment group; control group)
(8) Group Structure (Treatment group; control group)
(9) Materials (Treatment group; control group)
(10) Outcome Characteristics
(11) Effect Size Calculation.

When it was not possible to code a particular variable, the column(s) was(were) left blank; in the computer analysis, blanks were given the value of -9 to distinguish them from variables (if any) which were coded 0.

The variables coded are presented on the following pages. A column has been added to the right side of each page, noting the number of coding sheets (out of 341) which did not include information on the given variable. Some of the variables were of only incidental interest, but some were hoped eventually to yield interesting sub-analyses (e.g., the mortality of subjects: initial size minus final size of treatment and control groups).

Some variables which were not included on the coding sheet may have possibly yielded other relationships of interest--for example, whether the investigator (mostly in the case of dissertations) was the teacher for both the treatment and control groups, as was the case in several studies, or whether the teachers were unaffiliated with the designing of the study; whether the treatment and control groups were from the same school, different schools in the same district, or different districts; whether the same teachers taught both treatment and control groups; whether the study was conducted over different years on the same population, with the base year being the control condition and the later year being the experimental (as in the case of several studies).

Variables Included on the Coding Form

Card	Cols.	Variable Name		Data missing* (out of 341) No.	%
1	3-6	STUDY	Study identification code	0	0
	7-8	COMP	Comparison code	0	0
	9-10	OUTCOME	Outcome code	0	0
	11-14	YEAR	Year in which study was reported	0	0
	15	FORM	Form in which study was reported 1. Journal article 2. Book 3. Masters thesis 4. Doctoral thesis 5. Unpublished article 6. Conference paper	0	0
2	1-2	SAGE1	Mean age of students in treatment group	4	1%
	3-4	GRADE1	Modal grade of treatment group	3	1%
	5-7	IQ1	Average IQ of treatment group	122	37%
	8	SIQ1	Source of treatment group IQ 1. Stated 2. Inferred	127	37%
	9	HIQ1	Homogeneity of treatment group IQ 1. Homogeneous 2. Heterogeneous	135	40%
	10-12	SSEX1	Percent female in treatment group	260	76%
	13-15	SRAC1	Percent minority in treatment group	322	94%
	16	SPMIN1	Predominant minority in treatment group 1. Mexican 4. Native American 2. Other Hispanic 5. Black 3. Asian 6. Other	321	94%
	17-19	SPPMIN1	Percent predominant minority in treatment group	331	97%
	20	SES1	Mean socioeconomic status of treatment group 1. Low 2. Medium 3. High	252	74%
	21	HSES1	Homogeneity of treatment group SES 1. Homogeneous 2. Heterogeneous	255	75%

---

*"Missing" indicates that information on the specific variable could not be found in a report, or (as on Cards 6 through 9) that the coder had no basis for inferring that some feature either was or was not included.

Card	Cols.	Variable Name		Data missing (out of 341) No.	%
2	22	HAND1	Treatment group handicap, if any 1. Vision impaired 2. Hearing impaired 3. Learning disabled 4. Emotionally disturbed 5. Multiple handicaps 6. Other	(Deleted)	
	23	GROUP1	Treatment group tracking 1. Not grouped 2. Low track 3. Medium track 4. High track	316	93
	24-26	NSBEG1	Initial size of treatment group	188	55%
	27-29	NSEND1	Final size of treatment group	13	4%
	30	SIZ1	School size of treatment group 1. Less than 50 2. 50 to 199 3. 200 to 499 4. 500 to 999 5. 1000 to 2000 6. More than 2000	263	77%
	31	COMM1	Community type of treatment group 1. Urban 2. Rural 3. Suburban	112	33%
3		ON CARD 3, COLUMNS 1-31 CONTAIN THE SAME INFORMATION ON THE CONTROL GROUP THAT CARD 2 DOES ON THE TREATMENT GROUP. ON CARD 3, THE VARIABLE NAMES END WITH 2 INSTEAD OF 1 (e.g., COMM2).			
4	1-2	NTEACH1	Number of teachers in treatment group	39	11%
	3-4	TAGE1	Mean teacher age in treatment group	292	86%
	5-6	NEXP1	Treatment group teachers, average number of years of teaching	272	80%
	7-8	NSCI1	Average number of years of science teaching	302	89%
	9-10	NCURR1	Average number of years teaching this curriculum	328	96%
	11-13	TSEX1	Percent female teachers in treatment group	273	80%
	14-16	TRAC1	Percent minority teachers in treatment group	326	96%
	17	TPMIN1	Predominant minority of treatment group teachers 1. Mexican 2. Other Hispanic 3. Asian 4. Native American 5. Black 6. Other	334	98%

Card	Cols.	Variable Name		Data missing (out of 341) No.	%
4	18-20	TPPMIN1	Percent predominant minority teachers in treatment group	334	98%
	21	TBACK1	Educational background of treatment group teachers 1. Less than B.A. 2. B.A. only 3. B.A. + 15 units 4. M.A. only 5. M.A. + 15 units 6. M.A. + 30 units 7. Doctorate	291	85%
	22	TPSERV1	Treatment group teacher inservice training prior to experiment 1. Low; one-shot 2. Medium: series of lectures or workshops 3. Specialization	284	83%
	23	TNSF1	Training through N.S.F.? 1. Yes 2. No	303	89%
	24	TUNIV1	Training obtained at university? 1. Yes 2. No	288	84%
	25	TLOCAL1	Training obtained locally? 1. Yes 2. No	298	87%
	26	ACCEPT1	Treatment group teachers' acceptance of philosophy 1. Low 2. Medium 3. High	264	77%
	27	SASS1	Assignment of students to treatment group 1. Stratified random 2. Random 3. Matched 4. Intact random 5. Intact nonrandom 6. Self-selected	0	0%
	28	TASS1	Assignment of teachers to treatment group 1. Random 2. Nonrandom 3. Self-selected 4. Crossed 5. Matched	6	2%
	29	VALID1	Treatment group rated internal validity 1. Low (intact, highly dissimilar) 2. Medium (random or intact, some threat) 3. High (random, low mortality)	77	23%

Card	Cols.	Variable Name		Data missing (out of 341)	
				No.	%
4	30	UNIT1	Treatment group unit of analysis 1. Individual 2. Classroom subgroup 3. Classroom 4. School 5. Other	4	1%
	31	TYPE1	Type of study 1. Correlational 2. Quasi-experimental 3. Experimental	4	1%
5		ON CARD 5, COLUMNS 1-31 CONTAIN THE SAME INFORMATION ON THE CONTROL GROUP THAT CARD 4 DOES ON THE TREATMENT GROUP. ON CARD 5, THE VARIABLE NAMES END WITH 2 INSTEAD OF 1.			
6	1	SUBMA1	Subject matter in treatment group 1. General science 5. Earth science 2. Life science 6. Chemistry 3. Physical science 7. Physics 4. Biology 8. Other	0	0%
	2-3	DURATN1	Duration of treatment group program in weeks	13	4%
	4-5	WEEKS1	Time elapsed prior to testing, in weeks	18	5%
	6-8	TIME1	Minutes per week of treatment	44	13%
	9-10	FREQ1	Frequency of testing, times per month	323	95%
	11	FIDCUR1	Treatment group fidelity to curriculum 1. Low 2. Medium 3. High	288	84%
	12	FIDTRE1	Fidelity to treatment 1. Low 2. Medium 3. High	274	80%
	13	SUPINT1	Nature of implementation 1. Supplemental 2. Integral	16	5%
	14	BEHOBJ1	Behavioral objectives in treatment group 1. Used 2. Not used	69	20%
	15	SELPAC1	Self paced in treatment group 1. Used 2. Not used	2	1%
	16	IMFEED1	Immediate feedback in treatment group 1. Used 2. Not used	51	15%

Card	Cols.	Variable Name		Data missing (out of 341)	
				No.	%
6	17	DIATEST1	Diagnostic testing and prescription in treatment group 1. Used 2. Not used	54	16%
	18	CAI1	Computer assisted instruction in treatment group 1. Used 2. Not used	2	1%
	19	CMI1	Computer managed instruction in treatment group 1. Used 2. Not used	2	1%
	20	CSE1	Computer simulated experiments in treatment group 1. Used 2. Not used	2	1%
	21	TEAM1	Team teaching in treatment group 1. Used 2. Not used	2	1%
	22	TTUTOR1	Teacher as tutor in treatment group 1. Used 2. Not used	18	5%
	23	PTUTOR1	Pupil as tutor in treatment group 1. Used 2. Not used	8	2%
	24	INDINS1	Individualized instruction in treatment group 1. Used 2. Not used	2	1%
	25	UNITAPP1	Unit approach to instruction in treatment group 1. Used 2. Not used	39	11%
	26	DEPT1	Departmentalized elementary school in treatment group 1. Used 2. Not used	4	1%
	27	USESO1	Source papers in treatment group 1. Used 2. Not used	2	1%
	28	TRAD1	Traditional science classroom in treatment group 1. Used 2. Not used	2	1%
7		ON CARD 7, COLUMNS 1-28 CONTAIN THE SAME INFORMATION ON THE CONTROL GROUP THAT CARD 6 DOES ON THE TREATMENT GROUP.			

Card	Cols.	Variable Name		Data missing (out of 341) No.	%
8	1-2	CLASIZ1	Average class size in treatment group	39	11%
	3	FLEXMOD1	Flexible modular scheduling in treatment group 1. Used 2. Not used	2	1%
	4	LARGE1	Large group organization 1. Used 2. Not used	10	3%
	5	MEDGRP1	Normal class grouping in treatment group 1. Used 2. Not used	9	3%
	6	SMLGRP1	Small group organization 1. Used 2. Not used	50	15%
	7	SINGLE1	Group of 1 student 1. Used 2. Not used	59	17%
	8	LABACT1	Laboratory activities in treatment group 1. Used 2. Not used	65	19%
	9	DEMO1	Teacher demonstrations in trtmt grp 1. Used 2. Not used	147	43%
	10	STRLAB1	Student lab activities structured in treatment group 1. Used 2. Not used	125	37%
	11	UNSTR1	Student lab activities unstructured in treatment group 1. Used 2. Not used	150	44%
	12	MATER1	Nature of treatment group learning materials 1. Published 2. Modified published 3. Original	64	19%
	13	KITS1	Learning kits in treatment group 1. Used 2. Not used	109	32%
	14	LINPRO1	Linear programmed materials 1. Used 2. Not used	9	3%
	15	BRANCH1	Branched programmed materials 1. Used 2. Not used	9	3%

Card	Cols.	Variable Name		Data missing (out of 341) No.	%
8	16	GRREAD1	Programmed materials graded by reading level in treatment group 1. Used 2. Not used	14	4%
	17	SELFDIR1	Self-directed study 1. Used 2. Not used	6	2%
	18	STASS1	Student-assisted instructional program 1. Used 2. Not used	1	0%
	19	MEDBAS1	Media-based instruction 1. Television 2. Not used 3. Film 4. Teaching machines 5. Slides 6. Tapes	1	0%
	20	BBOARD1	Victor electrowriter 1. Used 2. Not used	2	1%
	21	MASTREQ1	Mastery learning 1. Required 2. Not required	151	44%
	22-24	MASTLEV1	Level of mastery required	335	98%
	25	TDIRREM1	Teacher-directed remediation 1. Used 2. t used	150	44%
	26	SDIRREM1	Student-directed remediation 1. Used 2. Not used	150	44%
	27	PSI1	Keller Personalized System of Instr. 1. Used 2. Not used	152	45%
	28	AUDTUT1	Audio-Tutorial 1. Used 2. Not used	159	47%
	29	CONTRAC1	Contracts for learning 1. Used 2. Not used	190	56%
9		ON CARD 9, COLUMNS 1-29 PROVIDE THE SAME INFORMATION ON THE CONTROL GROUP THAT CARD 8 DOES ON THE TREATMENT GROUP.			

Card	Cols.	Variable Name	
10	1-2	TYPCRIT	Type of outcome criterion 1. Cognitive low (recall, comprehension) 2. Cognitive high (application) 3. Cognitive mixed/general achievement 4. Problem solving 5. Affective toward subject 6. Affective toward science 7. Affective toward procedure/method 8. Values 9. Process skills 10. Methods of science 11. Psychomotor (lab skills) 12. Critical thinking 13. Creativity 14. Decision making 15. Logical thinking 16. Spatial reasoning 17. Self-concept 18. Science perceptions
	3	CONG1	Congruence of measure with treatment program 1. Low 2. Medium 3. High
	4	CONG2	Congruence of measure with control program 1. Low 2. Medium 3. High
	5	METHMS	Method of measurement (type of instrument) 1. Published, nationally available, standardized 2. Modification of national standardized 3. Ad hoc written tests 4. Classroom evaluation, excluding #1-3 5. Observation (passive, unstructured) 6. Structured interview, assessment 7. Other
	6	REACT	Reactivity of measure 1. Low; cognitive measures, 1 administration or long lag, not alterable 2. Medium 3. High; affective, transparent, alterable
	7-8	SOURCE	Calculation of effect size 1. Directly from reported or raw data 2. Reported with direct estimates (ANOVA, etc.) 3. From frequencies reported on ordinal scales 4. Backwards from other variances of means 5. Nonparametrics (other than #3) 6. Estimated from independent sources 7. Estimated from variance (correlation guessing) 8. Estimated from p-value 9. From raw data with teacher (year) effects removed 10. Other 11. From percentiles

Card	Cols.	Variable Name	
10	9	SOMEANS	Source of means 1. Unadjusted posttest 2. Covariance adjusted 3. Residual gains 4. Pre-post differences 5. Other
	10	SIGNIF	Reported significance 1. $p \leq .005$ 2. $.005 < p \leq .01$ 3. $.01 < p \leq .05$; 4. $.05 < p \leq .10$ 5. $p > .10$ 6. "not significant"
	11	DVUNITS	Dependent variable units 1. Grade-equivalent 2. Other
	12-15	GEU	Mean difference in grade equivalent units
	16	INDIV	Group variances reported individually 1. Yes 2. No
	17-20	RATIO	Ratio of treatment to control group standard deviation
	21-24	ESE	Effect size based on treatment group standard deviation
	25-28	ESC	Effect size based on control group standard deviation
	29-32	AVES	Average of ESE and ESC
	33-36	STES	Study Effect Size

SCIENCE PROJECT CODING SHEET - Ques. 2, Systems STANFORD

Author____
Title ____
____
Measure Used______________________________________

CARD 1
2 _
2 _ _ _
_ _
_ _
1 9 _ _
_

CARD 2
_ _
_ _
_ _ _
_
_
_ _ _
_ _ _
_
_ _ _
_
_
_
_
_ _ _
_ _ _
_
_

CARD 3
_ _
_ _
_ _ _
_
_
_ _ _
_ _ _
_
_ _ _
_
_
_
_
_ _ _
_ _ _
_
_

CARD 4
_ _
_ _
_ _
_ _
_ _
_ _ _
_ _ _
_
_ _ _
_
_
_
_
_
_
_
_
_
_
_

CARD 5
_ _
_ _
_ _
_ _
_ _
_ _ _
_ _ _
_
_ _ _
_
_
_
_
_
_
_
_
_
_
_

CARD 6
_
_ _
_ _
_ _ _
_ _
_
_
_
_
_
_
_
_
_
_
_
_
_
_
_
_
_
_

CARD 7
_
_ _
_ _
_ _ _
_ _
_
_
_
_
_
_
_
_
_
_
_
_
_
_
_
_
_
_

CARD 8
_ _
_
_
_
_
_
_
_
_
_
_
_
_
_
_
_
_
_
_
_
_ _ _
_
_
_
_
_
_

CARD 9
_ _
_
_
_
_
_
_
_
_
_
_
_
_
_
_
_
_
_
_
_
_ _ _
_
_
_
_
_
_

CARD 10
_ _
_
_
_
_
_ _
_
_
_
_ _ . _ _
_
_ _ . _ _
_ _ . _ _
_ _ . _ _
_ _ . _ _
_ _ . _ _

Variables Generated Prior to the Analysis

Prior to the initiation of the analysis itself, several new variables were created from existing variables. These newly created variables are listed below.

IMMEDES1	A variable to indicate whether the experimental group was evaluated within four weeks of the conclusion of the intervention or after that time. 1. Immediate evaluation. 2. Delayed evaluation.
IMMEDES2	Contains similar information as IMMEDES1 but pertaining to the control group.
ALLTIME1	A variable containing the total length of time in minutes of the experimental group program (i.e., the duration of the experimental group intervention).
ALLTIME2	Contains similar information as ALLTIME1 but pertaining to the control group.
VALDESN1	A variable to indicate the manner in which subjects were allocated to the experimental group. 1. Allocation by stratified random or random sampling 2. Allocation by matching subjects or by randomly allocating intact groups. 3. Nonrandom allocation.
VALDESN2	Contains similar information as VALDESN1 but pertaining to the control group.

Variables Recoded Prior to the Analysis

In several cases existing values of certain variables were modified and regrouped prior to the analysis. The new value labels are listed below.

METHMS Method of measurement (type of instrument)

1. Published, nationally available, standardized.
2. Modified national standardized and ad hoc written tests (previous values 2 and 3 taken together).
4. All other types of evaluation (previous values 4, 5, 6 and 7 taken together).

TYPCRIT Type of outcome criterion

1. All cognitive and problem-solving (previous values 1, 2, 3 and 4 taken together).
5. All affective (previous values 5, 6, 7 and 8 taken together.
10. Science methods (previous values 10 and 18 taken together).

(All other value labels remain the same.)

SOURCE Calculation of effect size

1. Directly from reported or raw data (same as previous value 1).
2. By direct calculation from reported statistics (previous values 2 and 9 taken together).
3. Less trustworthy methods of effect size estimation (previous values 3, 4, 5, 6, 7, 8, 10 and 11 taken together).

SECTION II. CODING THE DATA

CODING THE DATA

Sampling and Coding

Coding of studies began with microfilmed dissertations sent from Colorado whose titles implied that they would fall into the appropriate domain of inquiry.

Studies available through ERIC were identified directly by Colorado. Since dissertations are usually available in either microfilm or microfiche, this prior identification allowed the tedious task of going through thirty years of Dissertation Abstracts to be skipped by the coding center. Copies of abstracts of ERIC-available science studies facilitated the identification of studies in the system area. Five shipments of microfilms were coded; studies available on microfiche were obtained locally.

Scanning the bibliographies of each study gave a file of possible leads, including journal articles, books, dissertations, and conference papers. Following up these references frequently disclosed that they did not pertain to science instruction; or they were descriptive rather than experimental; or they involved college students as subjects, and so forth.

Educational journals were scanned, volume by volume, from 1950 (or later initial date in some instances) to the present time. Likely sounding titles in the tables of contents were followed up. These articles frequently were not relevant to the investigation; many described the same studies that had already been coded in dissertation form.

The following journals were examined during the above process:

American Biology Teacher
American Educational Review Journal
Audiovisual Communication Review
Bulletin, National Association of Secondary School Principals
California Journal of Educational Research
Journal of Chemical Education
Journal of Computer Based Education
Journal of Educational Psychology
Journal of Educational Research
Journal of Experimental Education
Journal of Programmed Instruction
Journal of Research on Science Teaching
Harvard Educational Review
School Science and Mathematics
Science Education
Science Teacher

Dissertations were the source of 58.5% of the included studies; journal articles, 31.5%; and unpublished studies, 10%. In many cases, it should be noted that a given study may have been reported several times (as a dissertation, one or more journal articles and a conference paper). When this occurred the most complete reported version of the study was used as the basis of the coding performed here.

Sampling Restrictions

Various restrictions and conventions were adopted to limit the range of the sample of studies coded in the meta-analysis.

1. Age of Subjects. The meta-analysis was limited to studies using students in grades K through 12. As in many other areas of educational research, studies in science education are often conducted on college students, who are most accessible to researchers. The college setting also provides some features that are not commonly found in elementary or secondary schools, such as computers, teaching assistants, and open laboratories. The data here then include very little on computer-managed or computer-assisted instruction, computer simulated experiments, audio-tutorial systems, or the Keller Personalized System of Instruction (designed for use at the college level).

2. Geography. The investigations carried out here were limited to studies reported in the United States. Doctoral dissertations were a major source of information, and American dissertations were the only ones readily available. Studies published in other countries, if included, would have been limited to those written in English and accessible in journal or book form, thus producing an incomplete international picture.

3. Control Group Instruction. Only studies which used a control group taught in the "traditional" or "conventional" classroom manner were included. This restriction eliminated studies (particularly where the dependent variable was student attitude toward science) which included some form of science instruction for the treatment group and no science instruction at all for the control group.

4. Year of Publication of Study. The year 1950 was designated as the earliest date of publication for included studies. It was expected that

the bulk of science studies would have been conducted from the late 1950s through the mid-1970s, the period of generous governmental funding of the sciences. Examining the dates of the coded studies confirms the validity of this expectation.

Of the more than 300 studies purporting to investigate "systems" which were considered and rejected, the following reasons for rejection were documented:

42% - subjects were college-aged

33% - incomplete data, such as means but no other information, only interview data or no data, or levels of significance with no indication of direction

17% - no control group

6% - control groups which were not taught "traditionally"--e.g., comparing two levels of individualized instruction

2% - subjects were teachers rather than students.

SECTION III. ANALYZING THE DATA

ANALYZING THE DATA

All Effect Sizes Over All Studies

Overall, a total of 341 effect sizes were generated in the Teaching Systems area of the current meta-analysis. The mean effect size produced over all systems was 0.103 with a standard deviation of 0.414, indicating that, on the average, an innovative teaching system in this sample can only expect to be one-tenth of a standard deviation better than traditional science teaching. Below, this mean effect size over all systems will be considered and discussed as a function of selected variables thought to be of interest.

Table 1
Mean Effect Size by Year of Publication

Year	No. of Δ	$\bar{\Delta}$	Standard Deviation
1950	2	0.250	0.014
1951	2	0.870	0.495
1952	1	1.050	0.000
1956	2	0.035	0.050
1957	2	0.025	0.050
1959	8	-0.194	0.334
1960	7	0.069	0.161
1961	29	0.015	0.464
1962	14	-0.062	0.377
1963	14	0.054	0.495
1964	9	0.207	0.248
1965	18	0.111	0.221
1966	19	0.036	0.259
1967	5	-0.176	0.286
1968	19	0.058	0.378
1969	21	0.097	0.251
1970	27	0.081	0.455
1971	54	0.190	0.456
1972	23	0.071	0.493
1973	19	0.007	0.330
1974	17	-0.015	0.463
1975	6	0.482	0.282
1976	4	0.443	0.245
1977	9	0.631	0.526
1978	6	0.098	0.233
1979	2	0.430	0.325
1980	2	0.000	0.000

In all, the 130 studies coded gave rise to 341 effect sizes distributed over the years 1950 through 1980, with the bulk of the effect sizes being obtained in the years 1961 through 1974. The minimum mean effect size for any given year was -0.194 with a standard deviation of 0.334, occurring in 1959 (based on 8 effect sizes), and the maximum mean effect size was obtained in 1951 with a value of 0.870 and a standard deviation of 0.495 (based on 2 effect sizes). No overall trend is evident from the data.

Table 2
Mean Effect Size by Form of Publication

Form	No. of Δ	$\bar{\Delta}$	Standard Deviation
Journal article	96	0.201	0.480
Dissertation	214	0.064	0.377
Unpublished paper	25	-0.034	0.360
Conference paper	6	0.508	0.172
ALL	341	0.103	0.414

Studies were reported as journal articles (producing 96 effect sizes), dissertations (producing 214 effect sizes), unpublished papers (producing 24 effect sizes), or conference papers (producing 6 effect sizes). The mean effect size over all systems derived from studies reported in journals was 0.201 with a standard deviation of 0.480, and the mean effect size derived from studies reported in dissertations was 0.064 with a standard deviation of 0.377, illustrating the selection bias noted earlier by Glass et al. (1981, Chapter 7). The mean effect size derived from studies reported as unpublished papers was -0.034 with a standard deviation of 0.36, and the and the mean effect size derived from studies reported at conferences was 0.508 with a standard deviation of 0.172.

Table 3
Mean Effect Size by Grade

Grade	No. of Δ	$\bar{\Delta}$	Standard Deviation
1	5	0.524	0.289
2	3	-0.253	0.280
3	7	0.050	0.479
4	10	-0.024	0.151
5	28	0.121	0.258
6	19	-0.074	0.435
7	28	0.086	0.293
8	25	0.315	0.491
9	31	0.115	0.263
10	63	0.099	0.406
11	76	0.152	0.420
12	43	0.008	0.548
missing	3		
ALL	341	0.103	0.414

In the current meta-analysis, effect sizes were obtained for studies which drew their subjects from grades 1 through 12. The minimum mean effect size obtained was -0.253 with a standard deviation of 0.280 (based on three effect sizes) obtained in grade 2, and the maximum mean effect size of 0.524 with a standard deviation of 0.289 (based on 5 effect sizes) obtained in grade 1. No obvious relationship between magnitude of mean effect size and grade is readily apparent in the data.

It should be noted here that, due to the constraints of educational practice, such a breakdown as is being attempted here on the basis of grade tends to subdivide the effect sizes obtained into subgroups differing also by curriculum area (i.e., students in grade 10 tend to study biology, students in grade 11 tend to study chemistry, etc.).

Table 4
Mean Effect Size by Assignment to Groups

Assignment to Groups	No. of Δ	$\bar{\Delta}$	Standard Deviation
Stratified random	38	0.010	0.390
Random	79	0.150	0.477
Matched	41	0.088	0.339
Intact random	91	0.206	0.428
Intact nonrandom	86	-0.003	0.362
Self-selected	6	0.142	0.215
ALL	341	0.103	0.414

In the current meta-analysis, an attempt was made to attach to each effect size a variable whose value described the method which was used to allocate subjects to either the experimental or control group. This variable was categorical in nature, and in all had six values, one of which was allocated to each effect size. In Table 4 are reported the mean effect sizes generated when the total mean effect size is broken down by the six values of this variable.

Table 5
Mean Effect Size by Subject Matter

Subject Matter	No. of Δ	$\bar{\Delta}$	Standard Deviation
General Science	100	0.090	0.315
Life Science	12	0.155	0.201
Physical Science	16	0.134	0.286
Biology	76	0.150	0.483
Earth Science	7	0.084	0.216
Chemistry	73	0.146	0.441
Physics	54	-0.014	0.508
Other	3	0.093	0.330
ALL	341	0.103	0.414

Table 6
Mean Effect Size by Type of Outcome Criterion

Type of Outcome Criterion	No. of Δ	$\bar{\Delta}$	Standard Deviation
Cognitive: low	61	0.050	0.461
Cognitive: high	11	0.[illegible]94	0.394
General achievement	165	[illegible]	0.[illegible]98
Problem solving	12	[illegible]	0.[illegible]54
Affective toward subject	13	[illegible].0[illegible]6	0.236
Affective toward science	22	0.015	0.333
Affective toward method	6	0.217	0.404
Affective toward studying	4	0.030	0.251
Process skills	3	-0.107	0.199
Methods of science	12	0.350	0.475
Psychomotor (lab skills)	6	0.892	0.684
Critical thinking	7	0.234	0.311
Creativity	4	0.430	0.457
Decision making	2	0.080	0.014
Logical thinking	3	0.403	0.280
Self-concept	3	0.317	0.100
Science perceptions	7	0.211	0.298
ALL	341	0.103	0.414

Table 7
Mean Effect Size by Origin of Instrument Used

Method of Measurement (Type of Instrument)	No. of Δ	$\bar{\Delta}$	Standard Deviation
Published nationally; standardized	173	0.045	0.387
Modification of national standardized	27	0.187	0.365
Ad hoc written tests	131	0.113	0.398
Classroom evaluation	6	1.028	0.511
Structured interview, assessment	2	0.720	0.453
Missing	2		
ALL	341	0.103	0.414

Table 8
Mean Effect Size by Method Used to Calculate Effect Size

Calculation of Effect Size	No. of Δ	$\bar{\Delta}$	Standard Deviation
Directly from reported or raw data (means and variances)	179	0.099	0.435
Reported with direct estimates (ANOVA, ANCOVA, t, F)	115	0.160	0.408
Directly from frequencies reported on ordinal scales (Probit, χ^2)	2	0.265	0.375
Backwards from other variances of means with random assignment	4	-0.160	0.145
Nonparametrics (other than #3)	8	-0.030	0.150
Guessed from independent sources	23	0.011	0.233
Estimated from variance (correlation guessing)	6	0.043	0.528
Estimated directly from p-value	2	-0.735	0.163
From percentiles	2	0.210	0.184
ALL	341	0.103	0.414

Table 9
Mean Effect Size by the Means Used in the Effect Size Calculation

Source of Means	No. of Δ	$\bar{\Delta}$	Standard Deviation
Unadjusted posttest	162	0.125	0.448
Covariance adjusted	67	0.086	0.387
Pre-post differences	93	0.087	0.382
Other	18	0.024	0.358
Missing	1		
ALL	341	0.103	0.414

Mean Effect Size, System by System

Tables 10a and 10b list the mean effect size obtained for each system and subsystem. In Table 10a the mean effect size on all outcome variables combined is presented; Table 10b shows the mean effect size on all outcome variables combined is presented; table 10b shows the mean effect size for each outcome variable (e.g. cognitive, affective, science methods, self-concept, etc.) for each system and subsystem.

Since there is some variation in the way the data is consolidated within the two tables, a description is needed at this point. In Table 10a, the row labelled "ALL" includes each of the 341 individual effect sizes otabined from the studies integrated. The effect sizes found in the remainder of Table 10a, however, total up to more than 341 and their wieghted average is not that given in the "ALL" row for two reasons. First of all, the table contains rows with data on various subsystems which, of course, duplicate the system data summarized in the line immediately above each group of subsystems. Systems for which subsytem information is given are computer-linked, median-based, and programmed instruction. Second, as noted previously, some effect sizes have been listed in more than one system in cases where the system evaluated in a given study met the definition presented earlier for more than one system. This duplicate listing occurred in 93 instances; essentially all of them are the result of an effect size being listed in both individualized instruction and one of several other systems.

In Table 10b, the "All Systems" row at the bottom of the table is the weighted average of the mean effect sizes for each of the above systems (information has not been duplicated, however, by inclusion of subsystem information in this weighted average). This table shows the mean effect size for each outcome variable for each system as well as an overall mean effect size on each outcome variable for all systems combined.

Table 10a

MEAN EFFECT SIZE, SYSTEM BY SYSTEM, ON ALL OUTCOME VARIABLES COMBINED

System	$\bar{\Delta}$	No. of Δ	s.d.	Max. Δ	Min. Δ
ALL	0.10	341	0.41	1.74	-0.87
Audio-Tutorial	0.17	7	0.27	0.52	-0.27
Computer Linked	0.13	14	0.58	1.45	-0.58
CAI	0.01	5	0.74	1.23	-0.58
CMI	0.05	8	0.22	0.53	-0.19
CSE	1.45	1	0	1.45	1.45
Contracts for Learning	0.47	12	0.61	1.74	-0.38
Dept. Elem. School	-0.09	3	0.17	0.08	-0.25
Individualized Inst.	0.17	131	0.46	1.74	-0.85
Mastery Learning	0.64	13	0.43	1.74	0.08
Media Based Instr.	-0.02	100	0.37	1.22	-0.87
TV	0.06	40	0.35	0.77	-0.87
Film	-0.07	58	0.38	1.22	-0.74
Slides	-0.47	1	0	-0.47	-0.47
Tapes	-0.27	1	0	-0.27	-0.27
PSI (Pers. Syst. Inst.)	0.60	15	0.42	1.74	0.08
Programmed Instr.	0.17	52	0.48	1.36	-0.82
Branched	0.21	5	0.80	1.23	-0.42
Linear	0.17	47	0.44	1.36	-0.82
Self-Directed	0.08	27	0.38	0.87	-0.58
Source Papers	0.14	13	0.21	0.48	-0.19
Student Assisted	0.09	6	0.17	0.34	-0.13
Team Teaching	0.06	41	0.38	1.36	-0.76

MEAN EFFECT SIZE ON EACH OUTCOME VARIABLE FOR EACH SYSTEM AND SUBSYSTEM

	Cognitive		Affective		Science Methods		Psychomotor		Critical Thinking		Logical Thinking		Creativity		Self-Concept	
System	$\bar{\Delta}$	n	$\bar{\Delta}$	n	$\bar{\Delta}$	n	$\bar{\Delta}$	n	$\bar{\Delta}$	n	$\bar{\Delta}$	n	$\bar{\Delta}$	n	$\bar{\Delta}$	n
Audio-Tutorial	.09	5	.33	1											A2 .12	1
Computer-Linked	.22	11	-.17	3												
CAI	.16	4	-.58	1												
CMI	.05	6	.04	2												
CSE	1.45	1														
Contracts for Learning	.22	5	.33	3	1.24	2			.53	2						
Dept. Elem. Sch.	-.09	3														
Individualized Instr.	.12	102	.16	10	.43	9	1.17	2	.33	4	.50	2	.50	2	.37	2
Mastery Learning	.50	8	.52	2	1.24	2			.89	1						
Media Based Instr.	-.03	75	-.10	16	.12	5	-.08	1	.16	2			.77	1		
T.V.	.02	33	-.12	1	.17	4			.15	1			.77	1		
Film	-.06	40	-.10	15	- .10	1	-.08	1	.17	1						
Slides	-.47	1														
Tapes	-.27	1														
P.S.I.	.49	7	.52	2	1.24	2			.89	1	.40	3				
Programmed Instr.	.17	51	.20	1												
Branched	.21	5														
Linear	.17	46	.20	1												
Self-Directed	-.12	16	-.10	3	-.11	1			.17	1	.40	3	.50	2	.42	1
Source Papers	.14	9	-.19	1	.25	3										
Student Assisted	.11	2	.17	2					.02	1			-.04	1		
Team Teaching	.09	31	-.12	7	.18	3										
SYSTEMS	.10	325	.04	51	.47	28	.75	3	.39	12	.43	8	.43	4	.39	4

Group Sizes, System by System

For each effect size in the current meta-analysis, the final sizes of the treatment and control groups were recorded.

Table 11
Final Size of Treatment Group Within Each System

	No. of Δs	Mean Δs	Maximum	Minimum
All Δs	341	122.9	>999	14
Audio-Tutorial	7	39.7	57	15
Computer Linked	14	111.6	232	24
CAI	5	38.2	58	24
CMI	8	167.9	232	24
CSE	1	29	29	29
Contracts for Learning	12	31.9	63	20
Departmentalized El. Sch.	3	284.3	646	70
Individualized Instruction	131	60.1	321	14
Mastery Learning	13	24.5	35	20
Media Based	100	229.2	>999	15
Television	40	242.2	>999	70
Film	58	227.9	919	22
Personalized System of Inst.	15	30.8	52	20
Programmed Instruction	52	73.7	186	18
Branched	5	39.0	58	26
Linear	47	77.7	186	18
Self-directed Study	27	51.3	122	23
Source Papers	13	35.7	50	25
Student Assisted	6	62.7	68	48
Team Teaching	41	100.3	261	25

Table 12
Final Size of Control Group Within Each System

	No. of △'s	Mean n	Maximum	Minimum
All	341	122.9	900	15
Audio-Tutorial	7	40.7	56	15
Computer Linked	14	87.9	233	20
CAI	5	34.2	52	20
CMI	8	127.5	233	23
CSE	1	39	39	39
Contracts for Learning	12	28.9	49	20
Departl. Elem. Sch.	3	356.7	707	175
Individualized Instr.	131	65.1	499	15
Mastery Learning	13	19.7	23	18
Media Based	100	181.1	900	17
Television	40	145.2	520	70
Film	58	212.4	900	17
PSI	15	26.7	51	18
Programmed Instr.	52	68.6	176	18
Branched	5	36.6	52	26
Linear	47	72.2	176	18
Self-directed Study	27	48.0	98	20
Source Papers	13	35.7	50	25
Student Assisted	6	48.3	64	25
Team Teaching	41	103.2	338	25

Summary Data for the Individual Systems

On the following pages, summary data are given and discussed for each of the teaching systems taken separately. The systems are arranged in alphabetical order, and in each case a summary data table is included.

Audio-Tutorial System. Seven effect sizes were obtained for the Audio-Tutorial system, with a mean effect size of 0.170 and a standard deviation of 0.274. The effect sizes were obtained from studies performed in the years 1970, 1972, 1974, and 1976, and although the mean effect sizes generated in each of these years vary considerably, definitive statements concerning their relative magnitude are difficult to make due to the small number of effect sizes concerned. However, the maximum mean effect size of 0.335 (based on 2 effect sizes) was obtained in 1976 and the minimum effect size of 0.000 (based on 1 effect size) was obtained in 1970.

When the effect sizes are broken down by form of reporting, the mean effect size obtained from journal articles is 0.223, with a standard deviation of 0.268 (based on 3 effect sizes) and the mean effect size from dissertations is 0.130 with a standard deviation of 0.312 (based on 4 effect sizes). Effect sizes were obtained in grades 3, 4, 9, and 10, the minimum being obtained in grade 3 with a mean effect size of 0.000 (based on 1 effect size), and a maximum in grade 10 with a mean effect size of 0.335 (based on 2 effect sizes). All effect sizes in the audio-tutorial system sample were produced by studies which used randomized allocation of subjects to groups, and hence no statements can be made concerning the variations of mean effect sizes over the variable VALDESN (a variable to measure the validity of the experimental design).

Systems using the audiotutorial technique were evaluated in two curriculum areas: general science and biology. The mean effect size obtained in biology was greater than the mean effect size obtained in general science with biology having a mean effect size of 0.230 (based on 5 effect sizes) and general science having a mean effect size of 0.020 (based on 2 effect sizes).

Effect sizes were obtained on various types of outcome criteria in studies that evaluated the audio-tutorial system. Effect sizes based on cognitive outcome criteria registered a mean effect size of 0.088 and a standard deviation of 0.287 (based on 5 effect sizes) and effect sizes based on affective outcome criteria registered a mean effect size of 0.330 (based on 1 effect size). The tests that were used to evaluate the effect size of the audio-tutorial system were published tests, modified published tests, and ad hoc tests produced by the investigator. Effect sizes generated by studies which made use of published test materials produced a mean effect size of 0.375 with a standard deviation of 0.064 (based on 2 effect sizes) whereas effect sizes produced by studies which made use of modified published tests and ad hoc tests taken together had a mean effect size of 0.088 and a standard deviation of 0.287 (based on 5 effect sizes).

Studies which reported the raw data in the account of the investigation were able to generate effect sizes directly from the raw data and such studies produced a mean effect size of 0.130, with a standard deviation of 0.312 (based on 4 effect sizes). Other studies reported the results of their investigation as a statistic or group of statistics (t, F, etc.) and mean effect sizes were calculated by methods due to Glass et al. (1981), producing a mean effect size of 0.335 with a standard deviation of 0.262 (based on 2 effect sizes).

Table 13
AUDIO-TUTORIAL SYSTEM

$\bar{\Delta}$ = 0.170
s.d. = 0.274
N = 7

	$\bar{\Delta}$	Standard Deviation	N
By Year of Publication			
1970	0.000	0.000	1
1972	0.040	0.000	1
1974	0.160	0.375	3
1976	0.335	0.262	2
By Form of Reporting			
Journal article	0.223	0.268	3
Dissertation	0.130	0.312	4
By Grade Level of Subjects			
3	0.000	0.000	1
4	0.040	0.000	1
9	0.160	0.375	3
10	0.335	0.262	2
By Validity of Design			
Random	0.170	0.274	7
By Subject Matter			
General Science	0.020	0.028	2
Biology	0.230	0.311	5

	$\bar{\Delta}$	Standard Deviation	N
By Immediate or Retention			
Immediate	0.223	0.268	3
Missing information			4
By Type of Outcome Criterion			
Cognitive	0.088	0.287	5
Affective	0.330	0.000	1
Self-Concept	0.420	0.000	1
By Method of Measurement (Instrument)			
Published	0.375	0.064	2
Modified published & Ad hoc	0.088	0.287	5
By Calculation of Effect Size			
From raw data	0.130	0.312	4
By direct calc.	0.335	0.262	2
Less trustworthy	0.000	0.000	1
By Source of Means			
Unadjusted posttest	0.040	0.000	1
Pre-post differences	0.230	0.311	5
Other	0.000	0.000	1

Computer-Linked Systems. Studies addressing the efficacy of computer-linked systems generated 14 effect sizes with a mean of 0.134 and standard deviation of 0.583. The studies which gave rise to these effect sizes were performed in the years 1965 (4 effect sizes), 1971 (5 effect sizes), 1972 (4 effect sizes), and 1975 (1 effect size); studies performed in and before 1971 yielded negative effect sizes and studies performed after and including 1972 yielded positive effect sizes. The maximum effect size was produced in 1972, and was 0.575 with a standard deviation of 0.943 (based on 4 effect sizes) and the minimum effect size was produced in 1971 and was -0.148 with a standard deviation of 0.243 (based on 5 effect sizes).

Studies reported in journals gave rise to a mean effect size of 1.340 with a standard deviation of 0.156 (based on 2 effect sizes) whereas studies reported in dissertations had a mean effect size of -0.121 with a standard deviation of 0.247 (based on 11 effect sizes). Studies were performed at grades 5, 10, 11, and 12, the majority being performed in the senior grades. It appears as though the mean effect size increases with grade; however, such trends can be regarded as having no significance due to the small size of the sample addressed.

Studies performed in order to evaluate the effect size of computer-linked systems made use of randomized allocation of subjects to groups, random allocation of intact groups, and nonrandomized allocation; those studies which made use of randomized allocation generated a mean effect size of 0.470 with a standard deviation of 1.009 (based on 4 effect sizes) whereas studies which used nonrandomized allocation produced a mean effect size of -0.053 with a standard deviation of 0.114 (based on 4 effect sizes). Studies were performed in the curriculum areas of general science, chemistry and physics; the mean effect size obtained for chemistry and physics were similar in magnitude (0.143 and 0.174) and substantially larger than that obtained in general science (0.020).

Of the 14 effect sizes obtained from studies evaluating this system, 11 were obtained by immediate assessment of students at the conclusion of the experimental and control group interventions and the remaining 3 effect sizes were intended to evaluate the retention effects of the interventions. The mean effect size for immediate assessment was 0.221 with a standard deviation of 0.626 (based on 11 effect sizes) and the mean effect size for evaluating retention effects was -0.183 with a standard deviation of 0.240 (based on 3 effect sizes), indicating that positive effects generated by the computer linked systems in students involved in the experimental intervention decayed with time relative to students in the control group.

Effect sizes based on cognitive outcome criteria produced a mean effect size of 0.216 with a standard deviation of 0.618 (based on 11 effect sizes) and effect sizes based on affective outcome criteria produced a mean effect size of -0.167 with a standard deviation of 0.359 (based on 3 effect sizes). Studies which made use of published test materials generated a mean effect size of -0.158 with a standard deviation of 0.256 (based on 5 effect sizes), and studies which made use of modified published test materials and ad hoc test materials taken together generated a mean effect size of 0.297 with a standard deviation of 0.661 (based on 9 effect sizes), showing that it was more likely for investigators who authored their own evaluation instruments to register a larger effect size.

In those cases in which effect sizes were able to be calculated from raw data reported in the studies themselves, a mean effect size of 0.149 with a standard deviation of 0.64 (based on 11 effect sizes) was produced, whereas effect sizes calculated from reported statistics produced a mean effect size of -0.145 with a standard deviation of 0.064 (based on 2 effect sizes).

Table 14
COMPUTER-LINKED SYSTEMS

$\bar{\Delta} = 0.134$
s.d. = 0.583
N = 14

	$\bar{\Delta}$	Standard Deviation	N
By Year of Publication			
1965	-0.053	0.114	4
1971	-0.148	0.243	5
1972	0.575	0.943	4
1975	0.530	0.000	1
By Form of Reporting			
Journal	1.340	0.156	2
Dissertation	-0.121	0.247	11
Conference Paper	0.530	0.000	1
By Grade Level of Subjects			
5	0.020	0.108	3
10	-0.053	0.114	4
11	0.143	0.941	3
12	0.400	0.841	4
By Validity of Design			
Random	0.470	1.009	4
Matched & Intact Random	0.035	0.367	6
Nonrandom	-0.053	0.114	4
By Subject Matter			
General Science	0.020	0.108	3
Chemistry	0.143	0.941	3
Physics	0.174	0.606	8

	$\bar{\Delta}$	Standard Deviation	N
By Immediate or Retention			
Immediate	0.221	0.626	11
Retention	-0.183	0.240	3
By Type of Outcome Criterion			
Cognitive	0.216	0.618	11
Affective	-0.167	0.359	3
By Method of Measurement (Instrument)			
Published	-0.158	0.256	5
Modified Published & Ad hoc	0.297	0.661	9
By Calculation of Effect Size			
From raw data	0.149	0.640	11
By direct calculatn.	-0.145	0.064	2
Less trustworthy	0.530	0.000	1
By Source of Means			
Unadjusted posttest	-0.174	0.269	8
Pre-posttest differences	0.548	0.731	5
Other	0.530	0.000	1

Computer Assisted Instruction. Five effect sizes with a mean of 0.010 and a standard deviation of 0.743 were obtained from studies which evaluated computer assisted instructional systems. These studies were performed in the years 1971 and 1972. The mean effect size for studies in the year 1971 was -0.400 with a standard deviation of 0.028 (based on 2 effect sizes) and the mean effect size produced by studies performed in the year 1972 was 0.283 with a standard deviation of 0.908 (based on 3 effect sizes). Effect sizes for the CAI system were reported both in journals (with a mean of 1.230) and dissertations (with a mean of -0.295). Students in grades 11 and 12 were used as subjects for CAI evaluation, and a mean effect size of 0.143 with a standard deviation of 0.941 (based on 3 effect sizes) was obtained for grade 11 subjects and a mean effect size of -0.190 with a standard deviation of 0.552 (based on 2 effect sizes) was obtained for grade 12.

Studies which made use of the randomized allocation of subjects to groups produced a larger mean effect size of 0.143 than studies which made use of matched subjects or the random allocation of intact groups (with a mean effect size of -0.190).

The two curriculum areas which were addressed in these studies were chemistry (with a mean effect size of 0.143) and physics (with a mean effect size of -0.190). Of the 5 effect sizes, four were generated by immediate evaluation of experimental effects, giving rise to a mean effect size of 0.118 with a standard deviation of 0.812, and one was generated by delayed evaluation (a retention effect) giving rise to an effect size of -0.420. Both cognitive and affective outcome criteria were evaluated, with cognitive measures producing a mean effect size of 0.158 with a standard deviation of 0.769 (based on 4 effect sizes), and affective measures producing a mean effect size of -0.580 (based on 1 effect size).

The mean effect size produced by studies which made use of published test materials was -0.580 and the mean effect size produced by studies which made use of modified published and ad hoc materials taken together was 0.158 with a standard deviation of 0.769 (based on 4 effect sizes).

Table 15
COMPUTER ASSISTED INSTRUCTION

$\bar{\Delta} = 0.010$
s.d. = 0.743
N = 5

	$\bar{\Delta}$	Standard Deviation	N
By Year of Publication			
1971	-0.400	0.028	2
1972	0.283	0.908	3
By Form of Reporting			
Journal article	1.230	0.000	1
Dissertation	-0.295	0.341	4
By Grade Level of Subjects			
11	0.143	0.941	3
12	-0.190	0.552	2
By Validity of Design			
Random	0.143	0.941	3
Matched & Intact Random	-0.190	0.552	2
By Subject Matter			
Chemistry	0.143	0.941	3
Physics	-0.190	0.552	2

	$\bar{\Delta}$	Standard Deviation	N
By Immediate or Retention			
Immediate	0.118	0.812	4
Retention	-0.420	0.000	1
By Type of Outcome Criterion			
Cognitive	0.158	0.769	4
Affective	-0.580	0.000	1
By Method of Measurement (Instrument)			
Published	-0.580	0.000	1
Modified published & Ad hoc	0.158	0.769	4
By Calculation of Effect Size			
From raw data	0.010	0.743	5
By Source of Means			
Unadjusted posttest	-0.295	0.341	4
Pre-post differences	1.230	0.000	1

Computer Managed Instruction. Teaching systems based on the use of computer managed instruction had a mean effect size of 0.048 with a standard deviation of 0.220 (based on 8 effect sizes). These studies were performed in the years 1965 (4 effect sizes), 1971 (3 effect sizes), and 1975 (1 effect size). The maximum effect size was obtained in 1975 and was 0.530, and the minimum mean effect size was obtained in 1965 and was -0.053. The mean effect size for studies reported as dissertations was -0.021 with a standard deviation of 0.109 (based on 7 effect sizes) and the mean effect size for studies reported at conferences was 0.530 (based on 1 effect size).

Grades 5, 10, and 12 were used to provide subjects for studies investigating computer managed instruction, the largest mean effect size being produced in grade 12, the minimum in grade 10; however, the small size of the sample of effect sizes being discussed here prevents definitive statements being made concerning any trend across grade level.

None of the studies being addressed here made use of the randomized allocation of subjects to groups; however, studies which made use of matched allocation and the random allocation of intact groups (taken together) produced a mean effect size of 0.148 with a standard deviation of 0.270 (based on 4 effect sizes) and studies which made use of nonrandomized allocation produced a mean effect size of -0.053 with a standard deviation of 0.114 (based on 4 effect sizes). The curriculum areas of general science and physics were the basis of the evaluation of the computer managed instruction system in the sample of studies being meta-analyzed here. In both cases, the mean effect sizes were positive but close to zero.

Of the 8 effect sizes in this subsection, 6 address the question of immediate effects, giving a mean effect size of 0.085 and 2 address the question of delayed effects, giving a mean effect size of -0.065. Those effect sizes for which cognitive outcome criteria were employed generated a mean effect size of 0.050 with a standard deviation of 0.260 (based on 6 effect sizes) and those effect sizes for which affective outcome criteria were employed generated a mean effect size of 0.040 with a standard deviation of 0.028 (based on 2 effect sizes). A mean effect size of -0.053 with a standard deviation of 0.114 (based on 4 effect sizes) was obtained in those cases in which published test materials were used, a mean effect size of 0.148 with a standard deviation of 0.27 (based on 4 effect sizes) was obtained in those cases in which modified published test materials and ad hoc test materials were employed. Effect sizes deriving from the meta-analysis of reported raw data produced a mean effect size of 0.028 with a standard deviation of 0.079 (based on 5 effect sizes), whereas effect sizes based on the transformation of reported statistics produced a mean effect size of -0.145 with a standard deviation of 0.064 (based on 2 effect sizes).

Table 16
COMPUTER MANAGED INSTRUCTION

$\bar{\Delta} = 0.048$
s.d. = 0.220
N = 8

	$\bar{\Delta}$	Standard Deviation	N
By Year of Publication			
1965	-0.053	0.114	4
1971	0.020	0.108	3
1975	0.530	0.000	1
By Form of Reporting			
Dissertation	-0.021	0.109	7
Conference Paper	0.530	0.000	1
By Grade Level of Subjects			
5	0.020	0.108	3
10	-0.053	0.114	4
12	0.530	0.000	1
By Validity of Design			
Matched & Intact Random	0.148	0.270	4
Nonrandom	-0.053	0.114	4
By Subject Matter			
General Science	0.020	0.108	3
Physics	0.064	0.279	5

	$\bar{\Delta}$	Standard Deviation	N
By Immediate or Retention			
Immediate	0.085	0.234	6
Retention	-0.065	0.177	2
By Type of Outcome Criterion			
Cognitive	0.050	0.260	6
Affective	0.040	0.028	2
By Method of Measurement (Instrument)			
Published	-0.053	0.114	4
Modified published & Ad hoc	0.148	0.270	4
By Calculation of Effect Size			
From raw data	0.028	0.079	5
By direct calc.	-0.145	0.064	2
Less trustworthy	0.530	0.000	1
By Source of Means			
Unadjusted posttest	-0.053	0.114	4
Pre-post differences	0.020	0.108	3
Other	0.530	0.000	1

Computer Simulated Experiments. In this meta-analysis, a single effect size of 1.450 was obtained for systems purporting to evaluate the use of computer simulated experiments, and as a consequence the breakdown of this effect size by other variables in the analysis was unable to be performed.

Table 17
COMPUTER SIMULATED EXPERIMENTS

$\bar{\Delta} = 1.450$
s.d. = 0.000
N = 1

Published 1972
Journal article
Grade 12
Random assignment
Physics
Immediate
Cognitive
Modified published & ad hoc
From raw data
Pre-post differences

Contracts for Learning. A mean effect size of 0.467 with a standard deviation of 0.605 (based on 12 effect sizes) was obtained for studies purporting to evaluate the use of contracts for learning. These studies were performed in the early 1970s and generated a maximum mean effect size of 0.857 with a standard deviation of 0.467 (based on 7 effect sizes) in 1971, and a minimum mean effect size of -0.255 with a standard deviation of 0.177 (based on 2 effect sizes) in 1974. The mean effect size obtained from studies reported in journals (0.610) was considerably higher than the mean effect size obtained from studies reported as dissertations (0.040). The contracts for learning system was evaluated in grades 8, 9, and 11, the maximum mean effect size being obtained in grade 8 (0.857) and the minimum mean effect size being obtained in grade 9 (-0.255).

Studies which made use of the randomized allocation of subjects to groups produced a mean effect size of 0.857 with a standard deviation of 0.467 (based on 7 effect sizes) and studies which made use of matched allocation and the random allocation of intact groups taken together produced a mean effect size of -0.078 with a standard deviation of 0.201 (based on 5 effect sizes). The curriculum areas of biology and chemistry were the areas that were used in the evaluation of this teaching system. Nine effect sizes with a mean of 0.610 and a standard deviation of 0.639 were obtained in the curriculum area of biology, and 3 effect sizes with a mean of 0.040 and a standard deviation of 0.114 were obtained in chemistry.

Of the 12 effect sizes seeking to address the effectiveness of the contracts for learning system, 11 concerned the evaluation of immediate effects (with a mean effect size of 0.522) and one sought to evaluate delayed effects (with an effect size of -0.130). Those effect sizes based on cognitive outcome criteria produced a mean effect size of 0.218 with a standard deviation of 0.569 (based on 5 effect sizes), and those based on affective outcome criteria produced a larger mean effect size of 0.330 with a standard deviation of 0.449 (based on 3 effect sizes). All effect sizes generated in this subsection originated in the use of published test materials as the basis of the outcome measures.

Effect sizes calculated on the basis of reported raw data amounted to five in all, and had a mean of -0.078 with a standard deviation of 0.201 and those calculated on the basis of the transformation of reported statistics had a mean of 0.857 with a standard deviation of 0.467 (7 effect sizes in all).

Table 18
CONTRACTS FOR LEARNING

$\bar{\Delta}$ = 0.467
s.d. = 0.605
N = 12

	$\bar{\Delta}$	Standard Deviation	N
By Year of Publication			
1971	0.857	0.467	7
1972	0.040	0.114	3
1974	-0.255	0.177	2
By Form of Reporting			
Journal article	0.610	0.639	9
Dissertation	0.040	0.114	3
By Grade Level of Subjects			
8	0.857	0.467	7
9	-0.255	0.177	2
11	0.040	0.114	3
By Validity of Design			
Random	0.857	0.467	7
Matched & Intact Random	-0.078	0.201	5
By Subject Matter			
Biology	0.610	0.639	9
Chemistry	0.040	0.114	3

	$\bar{\Delta}$	Standard Deviation	N
By Immediate or Retention			
Immediate	0.522	0.603	11
Retention	-0.130	0.000	1
By Type of Outcome Criterion			
Cognitive	0.218	0.569	5
Affective	0.330	0.449	3
Science Methods	1.235	0.714	2
Critical Thinking	0.530	0.509	2
By Method of Measurement (Instrument)			
Published	0.467	0.605	12
By Calculation of Effect Size			
From raw data	-0.078	0.201	5
By direct calculatn.	0.857	0.467	7
By Source of Means			
Unadjusted posttest	-0.078	0.201	5
Covariance adjusted	0.857	0.467	7

Departmentalized Elementary School. The departmentalized elementary school system (with a mean effect size of -0.090 and a standard deviation of 0.165 based on 3 effect sizes) was evaluated in studies reported in the years 1963, 1967, and 1969. The maximum effect size of 0.080 was obtained in 1963 and the minimum effect size of -0.250 was obtained in 1969. Effect sizes calculated from data reported in journal articles generated a mean of 0.080 whereas effect sizes calculated from data reported in dissertations had a lower mean of -0.175.

Those studies in which the raw data itself was reported produced a mean effect size of 0.080, while those studies which reported their outcomes as a statistic or group of statistics produced a mean effect size of -0.250.

Table 19
DEPARTMENTALIZED ELEMENTARY SCHOOL

$\bar{\Delta} = -0.090$
s.d. = 0.165
N = 3

	$\bar{\Delta}$	Standard Deviation	N
By Year of Publication			
1963	0.080	0.000	1
1967	-0.100	0.000	1
1969	-0.250	0.000	1
By Form of Reporting			
Journal article	0.080	0.000	1
Dissertation	-0.175	0.106	2
By Grade Level of Subjects			
5	-0.175	0.106	2
6	0.080	0.000	1
By Validity of Design			
Nonrandom	-0.090	0.165	3
By Subject Matter			
General Science	-0.090	0.165	3

	$\bar{\Delta}$	Standard Deviation	N
By Immediate or Retention			
Immediate	-0.090	0.165	3
By Type of Outcome Criterion			
Cognitive	-0.090	0.165	3
By Method of Measurement (Instrument)			
Published	-0.090	0.165	3
By Calculation of Effect Size			
From raw data	0.080	0.000	1
By direct calc.	-0.250	0.000	1
Less trustworthy	-0.100	0.000	1
By Source of Means			
Covariance adjusted	-0.010	0.127	2
Pre-post differences	-0.250	0.000	1

Individualized Instruction. Studies purporting to evaluate the use of systems based on the techniques of individualized instruction yielded 131 effect sizes. The mean effect size thus obtained was 0.174, with a standard deviation of 0.459. Studies addressing this question were reported between the years 1961 and 1980 with the majority of effect sizes being produced between 1969 and 1974. The maximum effect size of 0.806 was obtained in 1961 and the minimum effect size of -0.200 was obtained in 1967. The mean effect size for individualized instruction systems broken down by form of reporting illustrates a trend which is apparent in other facets of this meta-analysis, with the mean effect size for studies reported in journal articles (0.405) being considerably larger than the mean effect size for studies reported as dissertations (0.102).

The individualized instruction system was evaluated in grades 3 through 12, with the bulk of the evaluations being performed in grades 7 through 11; the mean effect sizes thus produced ranged from -0.100 to 0.467. Studies which made use of the randomized allocation of subjects to experimental groups produced a mean effect size of 0.215 with a standard deviation of 0.494 (based on 56 effect sizes), whereas studies which made use of matched allocation or the random allocation of intact groups produced a mean effect size of 0.175 with a standard deviation of 0.442 (based on 53 effect sizes). Studies which made use of the nonrandomized allocation of subjects to experimental groups produced a mean effect size of 0.070 with a standard deviation of 0.409 (based on 22 effect sizes). A trend of decreasing mean effect sizes is evident here as we move from true experimental designs through quasi-experimental designs of decreasing trustworthiness.

The individualized instruction system was evaluated across many curriculum areas, with the maximum mean effect size of 0.430 being obtained in the life science curriculum and the minimum effect size of 0.000 being obtained in the earth science curriculum, although in both of these areas only a single effect size was generated. In those areas in which the bulk of the effect sizes were produced (viz., general science with 36 effect sizes, biology with 30 effect sizes, and chemistry with 43 effect sizes), the maximum mean effect size of 0.265 with a standard deviation of 0.550 was obtained in biology and the minimum mean effect size of 0.016 with a standard deviation of 0.252 was obtained in general science.

Of the 120 effect sizes which were coded as appropriate to immediate or retention effects in this area (with 11 effect sizes having a missing value on this variable), 108 addressed the question of immediate effects and 12 addressed the question of retention effects. For those effect sizes which dealt with the immediate evaluation of experimental effects, the mean effect size was 0.22 with a standard deviation of 0.482, and for those effect sizes which dealt with the delayed evaluation of experimental effects, the mean effect size was -0.109 with a standard deviation of 0.234, illustrating that any difference in effect between experimental and control interventions decreased as time passed after the conclusion of the interventions. In fact, although individualized instruction is seen to be more effective than traditional instruction immediately on conclusion of the experimental treatment, once the treatment is withdrawn, and time has passed, traditional instruction is the system which retains its influence more effectively than individualized instruction.

The mean effect size generated by the individualized instruction system broken down by type of outcome criterion reveals a bewildering variability; however comparisons of mean effect sizes due to cognitive and affective outcome criteria indicate (given the extant standard deviations) little difference between the two. The mean effect size for those outcomes based on cognitive criteria is 0.118 with a standard deviation of 0.440 (based on 102 effect sizes), whereas the mean effect size for those outcomes based on affective criteria is 0.160 with a standard deviation of 0.373 (based on 10 effect sizes).

Those effect sizes which resulted from the use of published test materials revealed a mean which was of the same magnitude as the mean effect size obtained by use of modified published test materials and ad hoc test materials taken together (with the standard deviation in the two categories also being similar). The mean effect size attributed to the use of published test materials was 0.159 with a standard deviation of 0.442 (based on 65 effect sizes) and the mean effect size attributed to the use of modified published and ad hoc test materials being 0.159 with a standard deviation of 0.453 (based on 64 effect sizes). The mean effect size generated from those studies which reported raw data (0.176 with a standard deviation of 0.476 based on 72 effect sizes) was of a smaller magnitude than the effect size generated from those studies which required effect size calculation by transformation of common statistics. The mean effect size generated by such transformations was 0.236 with a standard deviation of 0.469 (based on 45 effect sizes).

Table 20
INDIVIDUALIZED INSTRUCTION

$\bar{\Delta}$ = 0.174
s.d. = 0.459
N = 131

	$\bar{\Delta}$	Standard Deviation	N
By Year of Publication			
1961	0.806	0.438	5
1963	0.190	0.651	2
1964	0.387	0.224	3
1965	0.047	0.133	3
1966	0.011	0.233	8
1967	-0.200	0.400	3
1968	0.076	0.397	9
1969	0.067	0.244	15
1970	0.047	0.392	19
1971	0.334	0.530	27
1972	0.316	0.639	10
1973	0.095	0.503	6
1974	-0.059	0.500	11
1975	0.507	0.376	3
1976	0.347	0.186	3
1977	0.210	0.184	2
1980	0.000	0.000	2
By Form of Reporting			
Journal article	0.405	0.519	29
Dissertation	0.102	0.422	100
Conference paper	0.450	0.113	2

	$\bar{\Delta}$	Standard Deviation	N
By Grade Level of Subjects			
3	0.000	0.000	1
4	-0.007	0.042	3
5	0.116	0.302	7
6	-0.100	0.461	8
7	0.027	0.226	17
8	0.404	0.585	15
9	0.192	0.328	14
10	0.112	0.440	20
11	0.215	0.493	40
12	0.467	0.723	6
By Validity of Design			
Random	0.215	0.494	56
Matched & Intact Random	0.175	0.442	53
Nonrandom	0.070	0.409	22
By Subject Matter			
General Science	0.016	0.252	36
Life Science	0.430	0.000	1
Physical Science	0.216	0.271	10
Biology	0.265	0.550	30
Earth Science	0.000	0.000	1
Chemistry	0.204	0.508	43
Physics	0.323	0.652	9
Other	0.030	0.000	1

(continued)

Individualized Instruction, continued

	$\bar{\Delta}$	Standard Deviation	N
By Immediate or Retention			
Immediate	0.220	0.482	108
Retention	-0.109	0.234	12
By Type of Outcome Criterion			
Cognitive	0.118	0.440	102
Affective	0.160	0.373	10
Science Methods	0.428	0.565	9
Psychomotor	1.165	0.064	2
Critical Thinking	0.325	0.405	4
Creativity	0.495	0.530	2
Self-Concept	0.365	0.078	2
By Method of Measurement (Instrument)			
Published	0.159	0.442	65
Modified published & Ad hoc	0.159	0.453	64
Other	1.165	0.064	2

	$\bar{\Delta}$	Standard Deviation	N
By Calculation of Effect Size			
From raw data	0.176	0.476	72
By direct calculation	0.236	0.469	45
Less trustworthy	-0.032	0.258	14
By Source of Means			
Unadjusted posttest	0.176	0.514	51
Covariance adjusted	0.198	0.467	32
Pre-post differences	0.150	0.412	40
Other	0.190	0.330	8

Mastery Learning. Thirteen effect sizes were obtained in the area of mastery learning; the mean effect size was 0.644 with a standard deviation of 0.430. Compared to other mean effect sizes reported in this meta-analysis, a mean effect size of 0.644 can be considered significant; however, as studies which purport to investigate the effects of mastery learning tended to remediate the experimental group on the basis of errors made by the participants on the outcome measure while the control group was not thus remediated, it is obvious that a large effect size should be obtained.

In this meta-analysis, studies investigating the mastery learning system were obtained for the years 1971 (7 effect sizes), 1975 (1 effect size), and 1977 (5 effect sizes). The maximum mean effect size of 0.857 with a standard deviation of 0.467 was obtained in 1971 and the minimum mean effect size of 0.368 with a standard deviation of 0.219 was obtained in 1977. Effect sizes originating from reports published in journals produced a mean effect size (0.713 with a standard deviation of 0.500) which was almost double the mean effect size (0.488 with a standard deviation of 0.161) reported at conferences.

The subjects which were used as participants in studies purporting to evaluate the effectiveness of mastery learning systems in this meta-analysis ranged across grades 8, 11, and 12; the maximum effect size of 0.857 was obtained in grade 8 and the minimum effect size of 0.368 was obtained in grade 11. The majority of the effect sizes included here originate from studies which utilized the random allocation of subjects to experimental groups, and such effect sizes have a mean of 0.742 with a standard deviation of 0.434 (based on 10 effect sizes). Studies which utilized the nonrandom allocation of subjects to experimental groups produced a mean effect size of 0.210 with a standard deviation of 0.184 (based on 2 effect sizes).

The curriculum areas addressed in this subsection are biology, chemistry, and physics. The maximum mean effect size of 0.857 was obtained in biology and the minimum mean effect size of 0.368 was obtained in chemistry. All effect sizes subsumed here are attributable to the immediate evaluation of experimental effects on conclusion of the experimental intervention.

The mean effect size due to cognitive outcome criteria was 0.498 with a standard deviation of 0.278 (based on 8 effect sizes) and the mean effect size due to affective outcome criteria was 0.515 with a standard deviation of 0.446 (based on 2 effect sizes). Those studies which made use of published test materials in their evaluation of the experimental and control groups produced a mean effect size of 0.713 with a standard deviation of 0.500 (based on 9 effect sizes) while those studies which utilized either modified published test materials or investigator-authored test materials produced a mean effect size of 0.488 with a standard deviation of 0.161 (based on 4 effect sizes). In those cases in which it was possible to calculate a mean effect size from raw data reported in the study itself, a mean effect size of 0.473 with a standard deviation of 0.194 (based on 3 effect sizes) was produced; studies which reported their outcomes as a statistic or group of statistics generated a mean effect size by transformation of those statistics of 0.857 with a standard deviation of 0.467 (based on 7 effect sizes).

Table 21
MASTERY LEARNING

$\bar{\Delta}$ = 0.644
s.d. = 0.430
N = 13

	$\bar{\Delta}$	Standard Deviation	N
By Year of Publication			
1971	0.857	0.467	7
1975	0.530	0.000	1
1977	0.368	0.219	5
By Form of Reporting			
Journal article	0.713	0.500	9
Conference paper	0.488	0.161	4
By Grade Level of Subjects			
8	0.857	0.467	7
11	0.368	0.219	5
12	0.530	0.000	1
By Validity of Design			
Random	0.742	0.434	10
Matched & Intact Random	0.530	0.000	1
Nonrandom	0.210	0.184	2
By Subject Matter			
Biology	0.857	0.467	7
Chemistry	0.368	0.219	5
Physics	0.530	0.000	1

	$\bar{\Delta}$	Standard Deviation	N
By Immediate or Retention			
Immediate	0.644	0.430	13
By Type of Outcome Criterion			
Cognitive	0.498	0.278	8
Affective	0.515	0.446	2
Science Methods	1.235	0.714	2
Critical Thinking	0.890	0.000	1
By Method of Measurement (Instrument)			
Published	0.713	0.500	9
Modified Published & Ad hoc	0.488	0.161	4
By Calculation of Effect Size			
From raw data	0.473	0.194	3
By direct calculation	0.857	0.467	7
Less trustworthy	0.317	0.226	3
By Source of Means			
Unadjusted posttest	0.368	0.219	5
Covariance-adjusted	0.857	0.467	7
Other	0.530	0.000	1

Media-Based Systems. Instructional systems based principally on the use of media (including film, television, and the like) as a means of inaugurating their effects gave rise to 100 effect sizes. The mean effect size in media-based systems was -0.023 with a standard deviation of 0.369. Within the media-based system, 40 effect sizes are attributable to television as a medium, and 58 are attributable to film as a medium; each of these will be reported in greater detail later. The mean effect size for television based systems was 0.055 with a standard deviation of 0.347, and the mean effect size for film based systems was -0.065 with a standard deviation of 0.379.

Studies of media-based instructional systems that were reported in the literature in the years from 1950 to 1973 are included here. The maximum mean effect size of 1.050 was derived from a study reported in 1952 and the minimum mean effect size of -0.558 was derived from studies reported in 1962. Although considerable variability exists in the mean effect sizes broken down by year of reporting, no overall trend is apparent. The mean effect size derived from studies reported in journals was -0.005 with a standard deviation of 0.393 (based on 37 effect sizes) whereas the mean effect size derived from studies reported as dissertations was -0.012 with a standard deviation of 0.370 (based on 47 effect sizes). Studies reported as unpublished articles produced a mean effect size of -0.097 with a standard deviation of 0.32 (based on 16 effect sizes). Although a slight downwards trend is evident here, the magnitude of the standard deviations involved prevents any such claim from being substantiated.

Subjects whose performances were evaluated in investigations of media-based systems were drawn from grades 1 through 12 with the bulk of the studies being performed at grades 5, 9, 11, and 12. The maximum mean effect size of 0.393 with a standard deviation of 0.012 (based on 3 effect sizes) was obtained at grade 1 and the minimum mean effect size of -0.262

with a standard deviation of 0.284 (based on 25 effect sizes) was obtained at grade 12; no clear relationship between mean effect size and grade is evident in the data.

Of the 100 effect sizes reported, 15 are attributable to studies which utilized a randomized allocation of subjects to groups, 41 are attributable to studies which utilized matched allocation of subjects or random allocation of intact groups, and 44 are attributable to studies which utilized nonrandom assignation. Studies which made use of the random allocation of subjects to groups produced a mean effect size of -0.219 with a standard deviation of 0.443, studies which made use of matched or random allocation of intact groups produced a mean effect size of 0.071 with a standard deviation of 0.266, and studies which used a nonrandomized allocation produced a mean effect size of -0.044 with a standard deviation of 0.402.

Effect sizes describing the media-based system were evolved from studies in several curriculum areas (general science, physical science, biology, chemistry, physics, and other). The maximum mean effect size of 0.149 with a standard deviation of 0.477 (based on 15 effect sizes) was obtained in the curriculum area of biology and the minimum effect size of -0.277 with a standard deviation of 0.288 (based on 27 effect sizes) was obtained in the area of physics.

Of the 97 effect sizes for which information concerning the timing of outcome measurement was available, 85 addressed the question of immediate effects and 12 addressed the question of delayed effects. For those effect sizes which were based on immediate evaluation of experimental effects, the mean effect size was -0.009 with a standard deviation of 0.377, and for those effect sizes which were based on the delayed evaluation of experimental effects, the mean effect size was -0.113, with a standard deviation of 0.347, illustrating an erosion of effect due to the media-based system relative to the effects created by traditional instruction over time.

In the case of effects evaluated as cognitive outcomes the mean effect size was -0.030 with a standard deviation of 0.388 (based on 75 effect sizes) and for the effects evaluated as affective outcomes the mean effect size was -0.104 with a standard deviation of 0.298 (based on 16 effect sizes). Effect sizes derived from studies making use of published test materials revealed a lower mean effect size (of -0.081) than effect sizes derived from studies making use of modified published test materials or investigator-authored test materials, these latter generating a mean effect size of 0.038. However, the magnitude of the standard deviations involved prevents substantiation of this claim. Effect sizes making use of published test materials make up 51% of the total and effect sizes making use of modified published or ad hoc test materials make up 49% of the total. Those effect sizes which were able to be calculated directly from raw data reported in the studies gave a mean effect size of -0.080 with a standard deviation of 0.345 (based on 42 effect sizes) and those effect sizes which were able to be calculated by transformation of reported statistics gave a mean effect size of 0.055 with a standard deviation of 0.413 (based on 44 effect sizes).

Table 22
MEDIA-BASED SYSTEMS

$\bar{\Delta} = -0.023$
s.d. = 0.369
N = 100

	$\bar{\Delta}$	Standard Deviation	N
By Type of Media			
Television	0.055	0.347	40
Film	-0.065	0.379	58
Slides	-0.470	0.000	1
Tapes	-0.270	0.000	1
By Year of Publication			
1950	0.250	0.014	2
1951	0.870	0.495	2
1952	1.050	0.000	1
1956	0.035	0.050	2
1957	0.025	0.050	2
1959	-0.194	0.334	8
1960	0.069	0.161	7
1961	-0.198	0.231	19
1962	-0.558	0.195	4
1963	-0.183	0.320	4
1964	0.117	0.221	6
1966	0.155	0.106	2
1968	0.208	0.310	6
1969	0.225	0.149	6
1970	-0.387	0.618	3
1971	0.028	0.366	19
1972	-0.124	0.249	5
1973	-0.130	0.481	2

	$\bar{\Delta}$	Standard Deviation	N
By Form of Reporting			
Journal article	-0.005	0.393	37
Dissertation	-0.012	0.370	47
Unpublished article	-0.097	0.320	16
By Grade Level of Subjects			
1	0.393	0.012	3
2	-0.253	0.280	3
3	0.058	0.524	6
4	-0.007	0.187	6
5	0.130	0.228	11
6	-0.190	0.561	7
7	0.180	0.000	2
8	0.045	0.120	2
9	0.116	0.160	11
10	0.390	0.445	6
11	0.007	0.327	17
12	-0.262	0.284	25
By Validity of Design			
Random	-0.219	0.443	15
Matched & Intact Random	0.071	0.266	41
Nonrandom	-0.044	0.402	44

(continued)

Media-Based Systems, continued

	$\bar{\Delta}$	Standard Deviation	N
By Subject Matter			
General Science	0.066	0.328	36
Physical Science	0.096	0.159	5
Biology	0.149	0.477	15
Chemistry	-0.009	0.324	15
Physics	-0.277	0.288	27
Other	0.125	0.460	2
By Immediate or Retention			
Immediate	-0.009	0.377	85
Retention	-0.133	0.347	12
By Type of Outcome Criterion			
Cognitive	-0.030	0.388	75
Affective	-0.104	0.298	16
Science Methods	0.118	0.143	5
Psychomotor	-0.080	0.000	1
Critical Thinking	0.160	0.014	2
Creativity	0.770	0.000	1
By Method of Measurement (Instrument)			
Published	-0.081	0.351	51
Modified published & ad hoc	0.038	0.381	49

	$\bar{\Delta}$	Standard Deviation	N
By Calculation of Effect Size			
From raw data	-0.080	0.345	42
By direct calculation	0.055	0.413	44
Less trustworthy	-0.097	0.244	14
By Source of Means			
Unadjusted posttest	-0.042	0.393	40
Covariance adjusted	-0.048	0.279	26
Pre-post differences	0.071	0.416	28
Other	-0.225	0.250	6

Television. Of the 100 effect sizes summarized previously under the heading of "Media-Based Instruction," 40 made use of television as the medium of instruction. It is these 40 effect sizes that will be dealt with here. Television-based instruction systems produced a mean effect size of 0.055 with a standard deviation of 0.347 and were reported in the years between 1957 and 1971. The mean effect size for studies reporting their outcomes in journals was 0.110 with a standard deviation of 0.194 (based on 10 effect sizes) and the mean effect size derived from studies reported as dissertations was 0.026 with a standard deviation of 0.411 (based on 26 effect sizes)--illustrating again the trend towards higher effects being reported in journals than in dissertations. Studies evaluating this system were performed at grades 1 through 9, and grade 12, with no substantial trend being apparent across the grades. Studies which made use of the randomized allocation of subjects to groups produced a mean effect size of 0.285 with a standard deviation of 0.686 (based on 2 effect sizes), the mean effect size derived from the matched allocation of subjects to groups or the random allocation of intact groups to treatment groups was 0.086 with a standard deviation of 0.287 (based on 34 effect sizes), and the mean effect size generated in studies which utilized the nonrandom allocation of subjects to experimental groups was -0.320 with a standard deviation of 0.522 (based on 4 effect sizes).

The curriculum areas of general science, physical science, biology, and physics formed the bodies of scientific expertise which were utilized in the evaluation of this system. Both general science and physical science generated positive but small mean effect sizes (0.092 and 0.096 respectively), whereas biology and physics generated negative mean effect sizes (-0.049 and -0.160 respectively). All effect sizes in this subsection derived from the immediate evaluation of experimental effects on the conclusion of the experiment.

The mean effect size for cognitive outcomes was 0.022 with a standard deviation of 0.355 (based on 33 effect sizes) and these effect sizes constituted the bulk of the effect sizes apparent in this subsection of the meta-analysis. In those studies which made use of published test materials, 7 effect sizes were generated with a mean effect size of 0.020 and a standard deviation of 0.119. All other studies in this area made use either of modified published test materials or investigator-authored test materials, and registered a mean effect size of 0.063 with a standard deviation of 0.379 (based on 33 effect sizes). Those effect sizes which were able to be calculated from raw data reported in the studies themselves produced a mean effect size of 0.018 with a standard deviation of 0.428 (12 effect sizes); of the remaining 28 effect sizes, 23 were produced by transformation of reported statistics and this group gave rise to a mean effect size of 0.066 with a standard deviation of 0.341.

Table 23
TELEVISION INSTRUCTION

$\bar{\Delta}$ = 0.055
s.d. = 0.347
N = 40

	$\bar{\Delta}$	Standard Deviation	N
By Year of Publication			
1957	0.060	0.000	1
1960	0.110	0.157	4
1964	0.090	0.115	3
1968	0.208	0.310	6
1969	0.205	0.188	4
1970	-0.387	0.618	3
1971	0.028	0.366	19
By Form of Reporting			
Journal article	0.110	0.194	10
Dissertation	0.026	0.411	26
Unpublished article	0.110	0.157	4
By Grade Level of Subjects			
1	0.393	0.012	3
2	-0.253	0.280	3
3	0.058	0.524	6
4	-0.007	0.187	6
5	0.197	0.135	7
6	-0.118	0.666	5
7	0.180	0.000	1
8	0.045	0.120	2
9	0.090	0.143	6
12	-0.120	0.000	1
By Validity of Design			
Random	0.285	0.686	2
Matched & Intact Random	0.086	0.287	34
Nonrandom	-0.320	0.522	4

	$\bar{\Delta}$	Standard Deviation	N
By Subject Matter			
General Science	0.092	0.342	26
Physical Science	0.096	0.159	5
Biology	-0.049	0.495	7
Physics	-0.160	0.057	2
By Immediate or Retention			
Immediate	0.055	0.347	40
By Type of Outcome Criterion			
Cognitive	0.022	0.355	33
Affective	-0.120	0.000	1
Science Methods	0.173	0.087	4
Critical Thinking	0.150	0.000	1
Creativity	0.770	0.000	1
By Method of Measurement (Instrument)			
Published	0.020	0.119	7
Modified published & Ad hoc	0.063	0.379	33
By Calculation of Effect Size			
From raw data	0.018	0.428	12
By direct calculation	0.066	0.341	23
Less trustworthy	0.096	0.159	5
By Source of Means			
Unadjusted posttest	0.018	0.428	12
Covariance adjusted	0.144	0.171	9
Pre-post differences	0.041	0.372	18
Other	-0.040	0.000	1

Film Based Instruction. Studies evaluated under the film based instruction subsection of the meta-analysis generated a total of 58 effect sizes with a mean effect size of -0.065 and a standard deviation of 0.378. The effect sizes were derived from studies reported between the years of 1950 and 1973 with the maximum mean effect size occurring in 1952 (1.050) and the minimum mean effect size occurring in 1962 (-0.558)--no obvious trend is apparent in the data. The mean effect size for studies reported in journals was -0.047 with a standard deviation of 0.440 (based on 27 effect sizes) whereas the mean effect size for studies reported in dissertations was -0.026 with a standard deviation of 0.311 (based on 19 effect sizes), reversing the trend apparent in other sections of the meta-analysis.

The subjects who formed the basis of the experimental and control groups in the evaluation of the film based instructional system were drawn from grades 5, 7, and 9 through 12, with the bulk of the effect sizes being obtained in grades 11 and 12. The minimum mean effect size of -0.268 with a standard deviation of 0.288 (based on 24 effect sizes) was obtained in grade 12 and the maximum mean effect size of 0.390 with a standard deviation of 0.445 (based on 6 effect sizes) was obtained in grade 10; no obvious trend is apparent in the data.

Studies whose groups were generated by random allocation of subjects produced a mean effect size of -0.283 with a standard deviation of 0.407 (based on 11 effect sizes) whereas those studies which utilized matched allocation of subjects or the random allocation of intact groups produced a mean effect size of 0.000 with a standard deviation of 0.107 (based on 7 effect sizes). The remaining 40 effect sizes were produced in studies which utilized a nonrandom allocation procudure and these effect sizes have a mean of -0.016 with a standard deviation of 0.385.

Curriculum areas addressed under the heading of film-based instruction were general science, biology, chemistry, physics, and other, with the minimum mean effect size of -0.287 with a standard deviation of 0.298 (based on 25 effect sizes) occurring in physics, and the maximum mean effect size of 0.323 with a standard deviation of 0.414 (based on 8 effect sizes) occurring in biology. Of the 58 effect sizes appertaining to film-based instructional systems, 55 possessed codings as to the immediate or delayed nature of their effects; the remaining 3 effect sizes were uncoded on this variable. Forty-three effect sizes were derived from the immediate evaluation of educational outcomes and these immediate effect sizes gave rise to a mean effect size of -0.051 with a standard deviation of 0.399, whereas the mean effect size based on delayed measurement of educational outcomes was -0.133 with a standard deviation of 0.347 (based on 12 effect sizes).

The mean effect size for film-based instructional systems based on cognitive outcome criteria was -0.055 with a standard deviation of 0.416 (based on 40 effect sizes) and the mean effect size based on affective outcome criteria was -0.103 with a standard deviation of 0.309 (based on 15 effect sizes). Those studies which made use of published test materials in their evaluation of treatment effects generated a mean effect size of -0.084 with a standard deviation of 0.377 (based on 42 effect sizes), while all other effect sizes in this subsection made use of either modified published test materials or investigator-authored test materials and generated a mean effect size of -0.014 with a standard deviation of 0.390. Of the 58 effect sizes, 28 were obtained from raw data contained in the studies themselves and the mean effect size thus obtained was -0.101 with a standard deviation of 0.308. The mean effect size obtained from the 21 effect sizes derived from studies which reported their outcomes as a statistic or group of statistics was 0.044 with a standard deviation of 0.488.

Table 24
FILM BASED INSTRUCTION

$\bar{\Delta}$ = -0.065
s.d. = 0.378
N = 58

	$\bar{\Delta}$	Standard Deviation	N
By Year of Publication			
1950	0.250	0.014	2
1951	0.870	0.495	2
1952	1.050	0.000	1
1956	0.035	0.050	2
1957	-0.010	0.000	1
1959	-0.194	0.334	8
1960	0.013	0.180	3
1961	-0.198	0.231	19
1962	-0.558	0.195	4
1963	-0.183	0.320	4
1964	0.143	0.327	3
1966	0.155	0.106	2
1969	0.265	0.021	2
1972	0.040	0.114	3
1973	-0.130	0.481	2
By Form of Reporting			
Journal article	-0.047	0.440	27
Dissertation	-0.026	0.311	19
Unpublished article	-0.166	0.335	12
By Grade Level of Subjects			
5	0.013	0.329	4
7	0.180	0.000	1
9	0.146	0.190	5
10	0.390	0.445	6
11	0.007	0.327	17
12	-0.268	0.288	24

	$\bar{\Delta}$	Standard Deviation	N
By Validity of Design			
Random	-0.283	0.407	11
Matched & Intact Random	0.000	0.107	7
Nonrandom	-0.016	0.385	40
By Subject Matter			
General Science	0.090	0.244	8
Biology	0.323	0.414	8
Chemistry	-0.009	0.324	15
Physics	-0.287	0.298	25
Other	0.125	0.460	2
By Immediate or Retention			
Immediate	-0.051	0.399	43
Retention	-0.133	0.347	12
By Type of Outcome Criterion			
Cognitive	-0.055	0.416	40
Affective	-0.103	0.309	15
Science Methods	-0.100	0.000	1
Psychomotor	-0.080	0.000	1
Critical Thinking	0.170	0.000	1
By Method of Measurement (Instrument)			
Published	-0.084	0.377	42
Modified published & Ad hoc	-0.014	0.390	16

(continued)

Film Based Instruction, continued

By Calculation of Effect Sizes	$\bar{\Delta}$	Standard Deviation	N
From raw data	-0.101	0.308	28
By direct calculation	0.044	0.488	21
Less trustworthy	-0.204	0.219	9

By Source of Means	$\bar{\Delta}$	Standard Deviation	N
Unadjusted posttest	-0.044	0.386	26
Covariance adjusted	-0.149	0.274	17
Pre-post differences	0.125	0.503	10
Other	-0.262	0.260	5

Personalized System of Instruction. The studies assessed in this meta-analysis which purported to evaluate the efficacy of the personalized system of instruction had a mean effect size of 0.603 with a standard deviation of 0.423 and there were 15 effect sizes in all. The studies appropriate to this system were reported in the years 1971, 1974, and 1977, and in each of the years the mean effect sizes were 0.857, 0.403, and 0.368 respectively.

Effect sizes generated from studies reported in journals gave rise to a mean effect size of 0.713 with a standard deviation of 0.500 (based on 9 effect sizes) and those derived from studies reported as unpublished articles gave rise to a mean effect size of 0.403 with a standard deviation of 0.280 (based on 3 effect sizes). The mean effect size derived from studies reported at conferences was 0.473 with a standard deviation of 0.194 (based on 3 effect sizes). This supports the trend evidenced earlier that mean effect sizes reported in journal articles tend, on the whole, to be larger than those reported elsewhere. Studies pertaining to the personalized system of instruction were carried out in grades 5, 8, and 11, and the maximum mean effect size of 0.857 was obtained in grade 8.

Studies which utilized the random allocation of subjects to experimental and control groups generated a mean effect size of 0.742 with a standard deviation of 0.434 (based on 10 effect sizes), studies which utilized the matched allocation or the random allocation of intact groups generated a mean effect size of 0.403 with a standard deviation of 0.280 (based on 3 effect sizes), and studies which made use of the nonrandom allocation of subjects to groups generated a mean effect size of 0.210 with a standard deviation of 0.184 (based on 2 effect sizes). The subject matter areas of general science, biology, and chemistry were used as curriculum areas in the evaluation of the personalized system of instruction, and the maximum effect size of 0.857 was obtained in biology. All 15 of the effect sizes in this subsection were generated by the evaluation of educational outcomes immediately on completion of the interventions.

The mean value of effect sizes generated by the use of cognitive outcome criteria was 0.493 with a standard deviation of 0.300 (based on 7 effect sizes) and the mean value of effect sizes generated by the use of affective outcome criteria was 0.515 with a standard deviation of 0.446 (based on 2 effect sizes).

Studies which made use of published test materials in their evaluation of experimental and control group effects produced a mean effect size of 0.713 with a standard deviation of 0.500 (based on 9 effect sizes) while studies which made use of modified published test materials or investigator-authored materials produced a mean effect size of 0.438 with a standard deviation of 0.219 (based on 6 effect sizes). In the case of the 6 effect sizes which were calculated directly from raw data reported in the studies, the mean value was 0.438 with a standard deviation of 0.219, whereas the mean value of the 7 effect sizes obtained by transformation of reported statistics was 0.857 with a standard deviation of 0.467.

Table 25
PERSONALIZED SYSTEM OF INSTRUCTION (PSI)

$\bar{\Delta} = 0.603$
s.d. = 0.423
N = 15

		Standard Deviation	
By Year of Publication			
1971	0.857	0.467	7
1974	0.403	0.280	3
1977	0.368	0.219	5
By Form of Reporting			
Journal article	0.713	0.500	9
Unpublished article	0.403	0.280	3
Conference paper	0.473	0.194	3
By Grade Level of Subjects			
5	0.403	0.280	3
8	0.857	0.467	7
11	0.368	0.219	5
By Validity of Design			
Random	0.742	0.434	10
Matched & Intact Random	0.403	0.280	3
Nonrandom	0.210	0.184	2
By Subject Matter			
General Science	0.403	0.280	3
Biology	0.857	0.467	7
Chemistry	0.368	0.219	5
By Immediate or Retention			
Immediate	0.603	0.423	15

		Standard Deviation	
By Type of Outcome Criterion			
Cognitive	0.493	0.300	7
Affective	0.515	0.446	2
Science Methods	1.235	0.714	2
Critical Thinking	0.890	0.000	1
Logical Thinking	0.403	0.280	3
By Method of Measurement (Instrument)			
Published	0.713	0.500	9
Modified Published & Ad hoc	0.438	0.219	6
By Calculation of Effect Size			
From raw data	0.438	0.219	6
By direct calculation	0.857	0.467	7
Less trustworthy	0.210	0.184	2
By Source of Means			
Unadjusted posttest	0.381	0.224	8
Covariance adjusted	0.857	0.467	7

Programmed Instruction. The 52 effect sizes which were collected under the umbrella of programmed instruction had a mean value of 0.174 with a standard deviation of 0.475. Studies appropriate to this area were reported in the years between 1961 and 1973 inclusive, and the mean effect sizes range from -0.200 in 1967 to 0.806 in 1961; no obvious trend is apparent in the data. The pattern recognized earlier concerning the mean effect sizes derived from studies reported in journals as opposed to studies reported in dissertations is repeated here. For effect sizes derived from journals, the mean effect size was .301 with a standard deviation of 0.448 (based on 7 effect sizes) while the mean effect size derived from studies reported in dissertations was 0.154 with a standard deviation of 0.480 (based on 45 effect sizes). Effect sizes were obtained from grades 4 and 6 through 12 with the maximum mean effect size of 1.07 occurring in grade 12 and the minimum mean effect size of -0.415 occurring in grade 8.

Studies which made use of random allocation of subjects to experimental and control groups produced a mean effect size of 0.173 with a standard deviation of 0.413 (based on 15 effect sizes), effect sizes derived from studies which made use of matched allocation or the random allocation of intact groups gave rise to a mean effect size of 0.186 with a standard deviation of 0.467 (based on 31 effect sizes), and studies which made use of nonrandom assignation gave rise to a mean effect size of 0.113, with a standard deviation of 0.710 (based on 6 effect sizes). General science, life science, physical science, biology, chemistry, and physics were the curriculum areas addressed under this system. The minimum mean effect size of -0.065 was obtained in general science and the maximum mean effect size of 0.533 was obtained in physics.

Of the 52 effect sizes reported in this area, 40 addressed the question of immediate experimental effects and these immediate effect sizes had a mean value of 0.260 with a standard deviation of 0.497. Eight effect sizes addressed the question of delayed experimental effects, and these effect sizes had a mean value of -0.113 with a standard deviation of 0.276, thus supporting the trend evidenced earlier in other subsections of the meta-analysis.

Studies which made use of published test materials gave rise to a mean effect size of 0.258 with a standard deviation of 0.394 (based on 10 effect sizes) whereas studies utilizing modified published test materials or investigator-authored test materials had a mean effect size of 0.154 with a standard deviation of 0.494 (based on 42 effect sizes). In those cases in which cognitive outcome criteria were used, the mean effect size was 0.173 with a standard deviation of 0.479 (based on 51 effect sizes). The mean effect size obtained from studies which reported their raw data was 0.173 with a standard deviation of 0.485 (based on 43 effect sizes) while the mean effect size obtained from those studies which reported their outcomes as one or more common statistics was 0.373 with a standard deviation of 0.420 (based on 6 effect sizes).

Table 26
PROGRAMMED INSTRUCTION

$\bar{\Delta}$ = 0.174
s.d. = 0.475
N = 52

	$\bar{\Delta}$	Standard Deviation	N
By Year of Publication			
1961	0.806	0.438	5
1963	0.190	0.651	2
1964	0.403	0.195	3
1965	0.047	0.133	3
1966	0.040	0.236	7
1967	-0.200	0.400	3
1968	0.088	0.494	6
1969	0.265	0.021	2
1970	-0.046	0.495	8
1971	0.013	0.310	7
1972	0.767	0.430	3
1973	0.173	0.780	3
By Form of Reporting			
Journal article	0.301	0.448	7
Dissertation	0.154	0.480	45
By Grade Level of Subjects			
4	-0.030	0.014	2
6	-0.070	0.521	7
7	0.023	0.342	4
8	-0.415	0.205	2
9	0.216	0.207	5
10	0.253	0.276	11
11	0.270	0.570	20
12	1.070	0.000	1
By Validity of Design			
Random	0.173	0.413	15
Matched & Intact Random	0.186	0.467	31
Nonrandom	0.113	0.710	6

	$\bar{\Delta}$	Standard Deviation	N
By Subject Matter			
General Science	-0.065	0.342	10
Life Science	0.430	0.000	1
Physical Science	0.148	0.161	4
Biology	0.055	0.424	12
Chemistry	0.291	0.550	22
Physics	0.533	0.516	3
By Immediate or Retention			
Immediate	0.260	0.497	40
Retention	-0.113	0.276	8
By Type of Outcome Criterion			
Cognitive	0.173	0.479	51
Affective	0.200	0.000	1
By Method of Measurement (Instrument)			
Published	0.258	0.394	10
Modified published & ad hoc	0.154	0.494	42
By Calculation of Effect Size			
From raw data	0.173	0.485	43
By direct calculation	0.373	0.420	6
Less trustworthy	-0.207	0.140	3
By Source of Means			
Unadjusted posttest	0.242	0.495	30
Covariance adjusted	-0.003	0.477	3
Pre-post differences	0.095	0.446	19

Branched Programmed Instruction. The branched programmed instructional system gave rise to 5 effect sizes with a mean effect size of 0.210 and a standard deviation of 0.798. The studies were reported in the years 1968, 1971, and 1972, with the maximum mean effect size of 1.230 being derived in the year 1972 and the minimum mean effect size of -0.400 being obtained in 1971. Again, we note that the mean effect size derived from journal entries (1.230) was larger than the mean effect size derived from studies reported in dissertations (-0.045).

Studies which made use of the random allocation of subjects to experimental and control groups produced a mean effect size of 0.143 with a standard deviation of 0.941 (based on 3 effect sizes), and the mean effect size obtained from studies which made use of either the matched allocation of subjects or the random allocation of intact groups of subjects was 0.310 with a standard deviation of 0.863 (based on 2 effect sizes). All effect sizes were obtained in the curriculum area of chemistry and were based on cognitive outcome measures, although 3 effect sizes were appropriate to the immediate evaluation of intervention effects and had a mean of 0.590 with a standard deviation of 0.854 and the remaining 2 effect sizes addressed the question of delayed effects and had a mean of -0.360 with a standard deviation of 0.085.

Table 27
BRANCHED PROGRAMMED INSTRUCTION

$\bar{\Delta}$ = 0.210
s.d. = 0.798
N = 5

	$\bar{\Delta}$	Standard Deviation	N
By Year of Publication			
1968	0.310	0.863	2
1971	-0.400	- 0.028	2
1972	1.230	0.000	1
By Form of Reporting			
Journal article	1.230	0.000	1
Dissertation	-0.045	0.645	4
By Grade Level of Subjects			
11	0.210	0.798	5
By Validity of Design			
Random	0.143	0.941	3
Matched & Intact Random	0.310	0.863	2
By Subject Matter			
Chemistry	0.210	0.798	5
By Immediate or Retention			
Immediate	0.590	0.854	3
Retention	-0.360	0.085	2
By Type of Outcome Criterion			
Cognitive	0.210	0.798	5

	$\bar{\Delta}$	Standard Deviation	N
By Method of Measurement (Instruments)			
Modified published & ad hoc	0.210	0.798	5
By Calculation of Effect Size			
From raw data	0.210	0.798	5
By Source of Means			
Unadjusted posttest	-0.400	0.028	2
Pre-post differences	0.617	0.809	3

Linear Programmed Instruction. Forty-seven effect sizes were obtained in this area and had a mean of 0.170 with a standard deviation of 0.441. The studies from which these effect sizes were drawn were reported between 1961 and 1973 and the mean effect sizes ranged in magnitude from -0.200 in 1967 to 0.806 in 1961, with no obvious trend being apparent in the data. Effect sizes derived from studies reported in journal articles had a mean effect size of 0.147 with a standard deviation of 0.199 (based on 6 effect sizes) whereas the mean effect size derived from studies reported in dissertations had a mean effect size of 0.173 with a standard deviation of 0.467 (based on 41 effect sizes). The samples of subjects which were used in the evaluation of linear programmed instructional systems were drawn from grades 4 and 6 through 12, with the minimum mean effect size of -0.415 being obtained in grade 8 and the maximum mean effect size of 1.070 being obtained in grade 12; no obvious trend is apparent in the data.

Studies which made use of the random allocation of subjects to groups generated a mean effect size of 0.180 with a standard deviation of 0.236 (based on 12 effect sizes), studies which made use of the matched allocation or the random allocation of intact groups generated a mean effect size of 0.178 with a standard deviation of 0.454 (based on 29 effect sizes) and studies which utilized a nonrandom assignation process generated a mean effect size of 0.113 with a standard deviation of 0.710 (based on 6 effect sizes).

In all, 6 separate curriculum areas were utilized in the evaluation of the effectiveness of linear programmed instructional systems, with the maximum mean effect size of 0.533 occurring in physics and the minimum mean effect size of -0.065 occurring in general science. Thirty-seven effect sizes addressed the question of immediate evaluation of experimental and control group effects, and had a mean effect size of 0.234 with a standard deviation of 0.467 and 6 effect sizes addressed the question of delayed intervention effects and had a mean effect size of -0.030 with a standard deviation of 0.269. Forty-six out of the 47 effect sizes made use of cognitive outcome criteria and had a mean effect size of 0.169 and a standard deviation of 0.446. The mean effect size obtained from studies which made use of published test materials was 0.258 with a standard deviation of 0.394 (based on 10 effect sizes) and the mean effect size obtained from studies which made use of either modified published test materials or investigator-authored test materials was 0.146 with a standard deviation of 0.455 (based on 37 effect sizes).

Table 28
LINEAR PROGRAMMED INSTRUCTION

$\bar{\Delta}$ = 0.170
s.d. = 0.441
N = 47

	$\bar{\Delta}$	Standard Deviation	N
By Year of Publication			
1961	0.806	0.438	5
1963	0.190	0.651	2
1964	0.403	0.195	3
1965	0.047	0.133	3
1966	-0.040	0.236	7
1967	-0.200	0.400	3
1968	-0.023	0.331	4
1969	0.265	0.021	2
1970	-0.046	0.495	8
1971	0.178	0.158	5
1972	0.535	0.219	2
1973	0.173	0.780	3
By Form of Reporting			
Journal article	0.147	0.199	6
Dissertation	0.173	0.467	41
By Grade Level of Subjects			
4	-0.030	0.014	2
6	-0.070	0.521	7
7	0.023	0.342	4
8	-0.415	0.205	2
9	0.216	0.207	5
10	0.253	0.276	11
11	0.290	0.508	15
12	1.070	0.000	1

	$\bar{\Delta}$	Standard Deviation	N
By Validity of Design			
Random	0.180	0.236	12
Matched & Intact Random	0.178	0.454	29
Nonrandom	0.113	0.710	6
By Subject Matter			
General Science	-0.065	0.342	10
Life Science	0.430	0.000	1
Physical Science	0.148	0.161	4
Biology	0.055	0.424	12
Chemistry	0.315	0.485	17
Physics	0.533	0.516	3
By Immediate or Retention			
Immediate	0.234	0.467	37
Retention	-0.030	0.269	6
By Type of Outcome Criterion			
Cognitive	0.169	0.446	46
Affective	0.200	0.000	1
By Method of Measurement (Instruments)			
Published	0.258	0.394	10
Modified published & ad hoc	0.146	0.455	37

(continued)

Linear Programmed Instruction, continued

From Calculation of Effect Size	$\bar{\Delta}$	Standard Deviation	N
From raw data	0.168	0.445	38
By direct calculation	0.373	0.420	6
Less trustworthy	-0.207	0.140	3

By Source of Means	$\bar{\Delta}$	Standard Deviation	N
Unadjusted posttest	0.288	0.480	28
Covariance adjusted	-0.003	0.477	3
Pre-post difference	-0.003	0.295	16

Self-Directed Study. Twenty-seven effect sizes were obtained from studies which purported to investigate the effects of self-directed study; these effect sizes had a mean value of 0.078 and a standard deviation of 0.375. The studies were reported in the years between 1968 and 1975, with the minimum mean effect size of -0.310 being derived from studies in 1971 and the maximum mean effect size originating from studies reported in 1975 (0.507). The trend concerning the relative magnitudes of mean effect sizes derived from studies reported in journals and dissertations, which has been referred to earlier in other areas of this meta-analysis, is again evidenced here; the mean effect size derived from studies reported in journals was 0.138 with a standard deviation of 0.544 (based on 4 effect sizes) whereas the mean effect size derived from studies reported in dissertations was -0.010 with a standard deviation of 0.328 (based on 19 effect sizes). Subjects who participated in the studies summarized here were drawn from grades 4, 5, 7, and 9 through 12, with the minimum mean effect size of -0.185 being obtained in grade 11 and the maximum mean effect size of 0.500 being obtained in grade 10; no obvious trend is apparent in the data.

Twelve of the 27 effect sizes considered here derived from studies which made use of the random allocation of subjects to experimental and control groups, and the mean effect size thus obtained was 0.107 with a standard deviation of 0.436. The remaining 15 effect sizes were obtained from studies which made use of either the matched allocation of subjects to experimental groups or the random allocation of intact groups, and the mean effect size in this case was 0.055 with a standard deviation of 0.334.

The curriculum areas of general science, biology, earth science, chemistry, and physics were used as content areas in the studies appropriate to this system. The minimum mean effect size of -0.047 was obtained in chemistry and the maximum mean effect size of 0.200 was obtained in general science. Twenty of the effect sizes collected here were intended to evaluate treatment effects immediately on conclusion of the intervention, and these effect sizes had a mean value of 0.095 with a standard deviation of 0.396. The mean effect size for delayed effects was -0.050 with a standard deviation of 0.523 (based on 2 effect sizes). Five effect sizes were uncoded on this variable.

In studies which made use of published test materials in order to evaluate the outcomes of the investigation, the mean effect size was 0.088 with a standard deviation of 0.392 (based on 16 effect sizes) while studies which made use of modified published test materials or investigator-authored test materials produced a mean effect size of 0.065 with a standard deviation of 0.368 (based on 11 effect sizes). The mean effect size for cognitive outcome criteria was -0.018 with a standard deviation of 0.341 (based on 16 effect sizes) and the mean effect size for affective outcome criteria was -0.097 with a standard deviation of 0.458 (based on 3 effect sizes). In those cases in which it was possible to generate effect sizes directly from reported raw data, the mean effect size thus obtained was 0.079 with a standard deviation of 0.348 (based on 20 effect sizes). in the case of effect sizes generated by transformation of reported statistics the mean effect size obtained was 0.495 with a standard deviation of 0.530 (based on 2 effect sizes).

Table 29
SELF-DIRECTED STUDY

$\bar{\Delta} = 0.078$
s.d. = 0.375
N = 27

	$\bar{\Delta}$	Standard Deviation	N
By Year of Publication			
1968	0.050	0.140	3
1970	0.505	0.262	2
1971	-0.310	0.208	4
1972	-0.067	0.271	7
1973	-0.275	0.106	2
1974	0.282	0.325	6
1975	0.507	0.376	3
By Form of Reporting			
Journal article	0.138	0.544	4
Dissertation	-0.010	0.328	19
Unpublished article	0.403	0.280	3
Conference paper	0.530	0.000	1
By Grade Level of Subjects			
4	0.040	0.000	1
5	0.403	0.280	3
7	0.050	0.140	3
9	0.008	0.371	5
10	0.500	0.342	4
11	-0.185	0.208	8
12	0.050	0.570	3
By Validity of Design			
Random	0.107	0.436	12
Matched & Intact Random	0.055	0.334	15

	$\bar{\Delta}$	Standard Deviation	N
By Subject Matter			
General Science	0.200	0.263	7
Biology	0.172	0.479	6
Earth Science	0.000	0.000	1
Chemistry	-0.047	0.355	10
Physics	0.050	0.570	3
By Immediate or Retention			
Immediate	0.095	0.396	20
Retention	-0.050	0.523	2
By Type of Outcome Criterion			
Cognitive	-0.018	0.341	16
Affective	-0.097	0.458	3
Science Methods	-0.110	0.000	1
Critical Thinking	0.170	0.000	1
Creativity	0.495	0.530	2
Logical Thinking	0.403	0.280	3
Self-Concept	0.420	0.000	1
By Method of Measurement (Instrument)			
Published	0.088	0.392	16
Modified published & ad hoc	0.065	0.368	11

(continued)

Self-Directed Study, continued

By Calculation of Effect Size	$\bar{\Delta}$	Standard Deviation	N
From raw data	0.079	0.348	20
By direct calculation	0.495	0.530	2
Less trustworthy	-0.092	0.386	5

By Source of Means	$\bar{\Delta}$	Standard Deviation	N
Unadjusted posttest	0.063	0.374	15
Covariance adjusted	-0.275	0.106	2
Pre-post differences	0.027	0.315	7
Other	0.507	0.376	3

Source Papers. Studies which purported to investigate the use of source papers as an instructional system yielded 13 effect sizes; the mean effect size was 0.142 with a standard deviation of 0.206. The studies concerned were reported in 1962 and 1966, with mean effect sizes of 0.136 and 0.163 respectively. All studies were reported as dissertations.

The study reported in 1962 drew its subjects from grade 10 and the curriculum area utilized was that of life science, whereas the study reported in 1966 drew its subjects from grade 7 and the curriculum area utilized was general science. All effect sizes addressed the question of intervention effects immediately on conclusion of the intervention. Where cognitive outcome criteria were utilized, the mean effect size was 0.142 with a standard deviation of 0.171 (based on 9 effect sizes), and where affective outcome criteria were utilized, the mean effect size was -0.190.

Table 30
SOURCE PAPERS

$\bar{\Delta}$ = 0.142
s.d. = 0.206
N = 13

	$\bar{\Delta}$	Standard Deviation	N
By Year of Publication			
1962	0.136	0.199	10
1966	0.163	0.274	3
By Form of Reporting			
Dissertation	0.142	0.206	13
By Grade Level of Subjects			
7	0.163	0.274	3
10	0.136	0.199	10
By Validity of Design			
Random	0.163	0.274	3
Matched & Intact Random	0.136	0.199	10
By Subject Matter			
General Science	0.163	0.274	3
Life Science	0.136	0.199	10

	$\bar{\Delta}$	Standard Deviation	N
By Immediate or Retention			
Immediate	0.142	0.206	13
By Type of Outcome Criterion			
Cognitive	0.142	0.171	9
Affective	-0.190	0.000	1
Science Methods	0.253	0.253	3
By Method of Measurement (Instrument)			
Published	0.183	0.220	10
Modified published & Ad hoc	0.007	0.006	3
By Calculation of Effect Size			
By direct calculation	0.163	0.274	3
Less trustworthy	0.136	0.199	10
By Source of Means			
Pre-post differences	0.142	0.206	13

Student Assisted Instructional System. The mean value of the 6 effect sizes obtained in the evaluation of this system was 0.088 with a standard deviation of 0.171, and all effect sizes were obtained in the year 1971. The mean effect size derived from studies reported as journal articles was 0.048 with a standard deviation of 0.205 (based on 4 effect sizes), whereas the mean effect size derived from studies reported as dissertations was 0.170 with a standard deviation of 0.014 (based on 2 effect sizes). All effect sizes were derived from studies utilizing the general science curriculum area and in all cases only the immediate evaluation of the intervention effects was addressed.

In the case of cognitive outcome criteria, the mean effect size was 0.105 with a standard deviation of 0.332 (based on 2 effect sizes), and in the case of affective outcome criteria, the mean effect size was 0.170 with a standard deviation of 0.014 (based on 2 effect sizes).

Table 31
STUDENT-ASSISTED INSTRUCTIONAL SYSTEM

$\bar{\Delta}$ = 0.088
s.d. = 0.171
N = 6

	$\bar{\Delta}$	Standard Deviation	N
By Date of Publication			
1971	0.088	0.171	6
By Form of Reporting			
Journal article	0.048	0.205	4
Dissertation	0.170	0.014	2
By Grade Level of Subjects			
5	0.048	0.205	4
6	0.170	0.014	2
By Validity of Design			
Random	0.048	0.205	4
Nonrandom	0.170	0.014	2
By Subject Matter			
General Science	0.088	0.171	6

	$\bar{\Delta}$	Standard Deviation	N
By Immediate or Retention			
Immediate	0.088	0.171	6
By Type of Outcome Criterion			
Cognitive	0.105	0.332	2
Affective	0.170	0.014	2
Critical Thinking	0.020	0.000	1
Creativity	-0.040	0.000	1
By Method of Measurement (Instrument)			
Published	-0.050	0.076	3
Modified published & Ad hoc	0.227	0.099	3
By Calculation of Effect Size			
By direct calculation	0.088	0.171	6
By Source of Means			
Unadjusted posttest	0.048	0.205	4
Pre-post differences	0.170	0.014	2

Team Teaching. Forty-one effect sizes were obtained in this subsection of the meta-analysis producing a mean effect size of 0.058 with a standard deviation of 0.378. Studies were reported between the years 1961 and 1980 with a large proportion being reported in 1962 and 1963; the minimum effect size of -0.365 was obtained in 1966 and the maximum mean effect size of 0.730 was obtained in 1976. The mean effect size obtained from studies reported in journals was 0.190 with a standard deviation of 0.357 (based on 8 effect sizes) and the mean effect size derived from studies reported in dissertations was 0.064, with a standard deviation of 0.347 (based on 26 effect sizes), supporting the trend noted earlier. The grade level of the subjects concerned ranged from grade 6 to grade 12, with the majority of effect sizes occurring in grade 10. The minimum mean effect size of -0.183 was obtained in grade 12, and the maximum mean effect size of 0.165 was was obtained in grade 7. Fourteen effect sizes were derived from studies which made use of the random allocation of subjects to experimental and control groups, and the mean effect size in this case was -0.004 with a standard deviation of 0.492. In the case of matched allocation to groups or random assignation of intact groups, the mean effect size was 0.161 with a standard deviation of 0.313 (based on 19 effect sizes). In those studies in which nonrandom allocation was utilized, the mean effect size was -0.076 with a standard deviation of 0.238 (based on 8 effect sizes).

In all, six different curriculum areas provided the underlying content basis for the evaluation of the team teaching system, with the minimum mean effect size of -0.490 occurring in physical science and the maximum mean effect size of 0.295 occurring in general science. Thirty-seven of the effect sizes addressed the question of immediate evaluation of intervention effects and the mean effect size in this case was 0.063. The mean effect size in the case of retention effects was 0.035 with a standard deviation of 0.007 (based on 2 effect sizes). Where cognitive outcome criteria were used, 31 effect sizes gave rise to a mean effect size of 0.087 with a standard deviation of 0.409, and where affective outcome criteria were used, a mean effect size of -0.124 with a standard deviation of 0.235 (based on 7 cases) was registered.

Twenty-three of the effect sizes owed their origin to the use of published test materials, and generated a mean effect size of 0.094 with a standard deviation of 0.394. The mean effect size derived from studies which made use of modified published test materials or investigator-authored test materials was 0.014, with a standard deviation of 0.361 (based on 16 effect sizes). In the case of the 17 effect sizes generated by studies which reported raw data, the mean effect size produced was -0.101 with a standard deviation of 0.374. In the case of the 9 effect sizes that were produced by the transformation of reported statistics, the mean effect size was 0.253, with a standard deviation of 0.479.

Table 32
TEAM TEACHING

$\bar{\Delta}$ = 0.058
s.d. = 0.378
N = 41

	$\bar{\Delta}$	Standard Deviation	N
By Date of Publication			
1961	0.032	0.273	5
1962	0.136	0.199	10
1963	0.026	0.521	11
1965	0.188	0.354	5
1966	-0.365	0.021	2
1968	-0.208	0.369	4
1969	0.470	0.000	1
1976	0.730	0.000	1
1980	0.030	0.000	2
By Form of Reporting			
Journal article	0.190	0.357	8
Dissertation	0.064	0.347	26
Unpublished article	-0.255	0.355	6
Conference paper	0.730	0.000	1
By Grade Level of Subjects			
6	0.000	0.000	2
7	0.165	0.926	2
9	0.030	0.022	4
10	0.123	0.430	21
11	0.013	0.204	6
12	-0.183	0.320	4
By Validity of Design			
Random	-0.001	0.492	14
Matched & Intact Random	0.161	0.313	19
Nonrandom	-0.076	0.238	8

	$\bar{\Delta}$	Standard Deviation	N
By Subject Matter			
General Science	0.295	0.389	4
Life Science	0.136	0.199	10
Physical Science	-0.490	0.000	1
Biology	0.062	0.487	16
Chemistry	-0.027	0.286	3
Physics	-0.081	0.270	7
By Immediate or Retention			
Immediate	0.063	0.398	37
Retention	0.035	0.007	2
By Type of Outcome Criterion			
Cognitive	0.087	0.409	31
Affective	-0.124	0.235	7
Science Methods	0.183	0.177	3
By Method of Measurement (Instrument)			
Published	0.094	0.394	23
Modified published & ad hoc	0.014	0.361	16
By Calculation of Effect Size			
From raw data	-0.101	0.374	17
By direct calculation	0.253	0.479	9
Less trustworthy	0.122	0.239	15
By Source of Means			
Unadjusted posttest	0.083	0.507	17
Covariance adjusted	0.060	0.095	3
Pre-post differences	0.021	0.270	16
Other	0.094	0.330	5

Mean Effect Sizes Broken Down by Selected Variables Across Systems

On the following pages, the mean effect sizes for all systems broken down by variables thought to be of interest are listed in order to facilitate inter-system comparison.

Table 33
Effect Sizes by Form of Reporting for Each System

	$\bar{\Delta}$ s.d. (N)	Journal Articles	Dissertations	Unpublished Articles	Conference Papers
All Δs	0.103 0.414 (341)	0.201 0.480 (96)	0.064 0.377 (214)	-0.034 0.360 (25)	0.508 0.172 (6)
Audio-Tutorial	0.170 0.274 (7)	0.223 0.268 (3)	0.130 0.312 (4)		
Computer Linked	0.134 0.583 (14)	1.340 0.156 (2)	-0.121 0.247 (11)		0.530 0.000 (1)
CAI	0.010 0.743 (5)	1.230 0.000 (1)	-0.295 0.341 (4)		
CMI	0.048 0.220 (8)		-0.021 0.109 (7)		0.530 0.000 (1)
CSE	1.450 0.000 (1)	1.450 0.000 (1)			
Contracts	0.467 0.605 (12)	0.610 0.639 (9)	0.040 0.114 (3)		
Dept. Elem. School	-0.090 0.165 (3)	0.080 0.000 (1)	-0.175 0.106 (2)		
Indiv. Instr.	0.174 0.459 (131)	0.405 0.519 (29)	0.102 0.422 (100)		0.450 0.113 (2)
Mastery Learning	0.644 0.430 (13)	0.713 0.500 (9)			0.488 0.161 (4)

Table 33, continued

	$\bar{\Delta}$ s.d. (N)	Journal Articles	Dissertations	Unpublished Articles	Conference Papers
All Δs	0.103 0.414 (341)	0.201 0.480 (96)	0.064 0.377 (214)	-0.034 0.360 (25)	0.508 0.172 (6)
Media Based	-0.023 0.369 (100)	-0.005 0.393 (37)	-0.012 0.370 (47)	-0.097 0.320 (16)	
TV	0.055 0.347 (40)	0.110 0.194 (10)	0.026 0.411 (26)	0.110 0.157 (4)	
Film	-0.065 0.378 (58)	-0.047 0.440 (27)	-0.026 0.311 (19)	-0.166 0.335 (12)	
PSI	0.603 0.423 (15)	0.713 0.500 (9)		0.403 0.280 (3)	0.473 0.194 (3)
Prog. Instr.	0.174 0.475 (52)	0.301 0.448 (7)	0.154 0.480 (45)		
Branched	0.210 0.798 (5)	1.230 0.000 (1)	-0.045 0.645 (4)		
Linear	0.170 0.441 (47)	0.147 0.199 (6)	0.173 0.467 (41)		
Self-Directed	0.078 0.375 (27)	0.138 0.544 (4)	-0.010 0.328 (19)	0.403 0.280 (3)	0.530 0.000 (1)
Source Papers	0.142 0.206 (13)		0.142 0.206 (13)		
Student Assisted	0.088 0.171 (6)	0.048 0.205 (4)	0.170 0.014 (2)		
Team Teaching	0.058 0.378 (41)	0.190 0.357 (8)	0.064 0.347 (26)	-0.255 0.355 (6)	0.730 0.000 (1)

Table 34
Effect Sizes by Grade Level of Subjects for Each System

	$\bar{\Delta}$ s.d. (N)	1	2	3	4	5	6	7	8	9	10	11	12	?
All Δs	0.103 0.414 (341)	0.524 0.289 (5)	-0.253 0.280 (3)	0.050 0.479 (7)	-0.024 0.151 (10)	0.121 0.258 (28)	-0.074 0.435 (19)	0.086 0.293 (28)	0.315 0.491 (25)	0.115 0.263 (31)	0.099 0.406 (63)	0.152 0.420 (76)	0.008 0.548 (43)	(3)
Audio-Tutorial	0.170 0.274 (7)			0.000 0.000 (1)	0.040 0.000 (1)					0.160 0.375 (3)	0.335 0.262 (2)			
Computer Linked	0.134 0.583 (14)					0.020 0.108 (3)					-0.053 0.114 (4)	0.143 0.941 (3)	0.400 0.841 (4)	
CAI	0.010 0.743 (5)											0.143 0.941 (3)	-0.190 0.552 (2)	
CMI	0.048 0.220 (8)					0.020 0.108 (3)					-0.053 0.114 (4)		0.530 0.000 (1)	
CSE	1.450 0.000 (1)												1.450 0.000 (1)	
Contracts	0.467 0.605 (12)								0.857 0.467 (7)	-0.255 0.177 (2)		0.040 0.114 (3)		
Dept. Elem. School	-0.090 0.165 (3)					-0.175 0.106 (2)	0.080 0.000 (1)							
Indiv. Instr.	0.174 0.459 (131)			0.000 0.000 (1)	-0.007 0.042 (3)	0.116 0.302 (7)	-0.100 0.461 (8)	0.027 0.226 (17)	0.404 0.585 (15)	0.192 0.328 (14)	0.112 0.440 (20)	0.215 0.493 (40)	0.467 0.723 (6)	
Mastery Learning	0.644 0.430 (13)								0.857 0.467 (7)			0.368 0.219 (5)	0.530 0.000 (1)	

(continued)

Table 34, continued

	$\bar{\Delta}$ s.d. (N)	1	2	3	4	5	6	7	8	9	10	11	12	?
All Δs	0.103 0.414 (341)	0.524 0.289 (5)	-0.253 0.280 (3)	0.050 0.479 (7)	-0.024 0.151 (10)	0.121 0.258 (28)	-0.074 0.435 (19)	0.086 0.293 (28)	0.315 0.491 (25)	0.115 0.263 (31)	0.099 0.406 (63)	0.152 0.420 (76)	0.008 0.548 (43)	(3)
Media Based	-0.023 0.369 (100)	0.393 0.012 (3)	-0.253 0.280 (3)	0.058 0.524 (6)	-0.007 0.187 (6)	0.130 0.228 (11)	-0.190 0.561 (7)	0.180 0.000 (2)	0.045 0.120 (2)	0.116 0.160 (11)	0.390 0.445 (6)	0.007 0.327 (17)	-0.262 0.284 (25)	
TV	0.055 0.347 (40)	0.393 0.012 (3)	-0.253 0.280 (3)	0.058 0.524 (6)	-0.007 0.187 (6)	0.197 0.135 (7)	-0.118 0.666 (5)	0.180 0.000 (1)	0.045 0.120 (2)	0.090 0.143 (6)			-0.120 0.000 (1)	
Film	-0.065 0.378 (58)					0.013 0.329 (4)		0.180 0.000 (1)		0.146 0.190 (5)	0.390 0.445 (6)	0.007 0.327 (17)	-0.268 0.288 (24)	
PSI	0.603 0.423 (15)					0.403 0.280 (3)			0.857 0.467 (7)			0.368 0.219 (5)		
Prog. Instr.	0.174 0.475 (52)				-0.030 0.014 (2)		-0.070 0.521 (7)	0.023 0.342 (4)	-0.415 0.205 (2)	0.216 0.207 (5)	0.253 0.276 (11)	0.270 0.570 (20)	1.070 0.000 (1)	
Branched	0.210 0.798 (5)											0.210 0.798 (5)		
Linear	0.170 0.441 (47)				-0.030 0.014 (2)		-0.070 0.521 (7)	0.023 0.342 (4)	-0.415 0.205 (2)	0.216 0.207 (5)	0.253 0.276 (11)	0.290 0.508 (15)	1.070 0.000 (1)	
Self-Directed	0.078 0.375 (27)				0.040 0.000 (1)	0.403 0.280 (3)		0.050 0.140 (3)		0.008 0.371 (5)	0.500 0.342 (4)	-0.185 0.208 (8)	0.050 0.570 (3)	
Source Papers	0.142 0.206 (13)							0.163 0.274 (3)			0.136 0.199 (10)			
Student Assisted	0.088 0.171 (6)					0.048 0.205 (4)	0.170 0.014 (2)							
Team Teaching	0.058 0.378 (41)						0.000 0.000 (2)	0.165 0.926 (2)		0.030 0.022 (4)	0.123 0.430 (21)	0.013 0.204 (6)	-0.183 0.320 (4)	

Table 35
Effect Sizes by Validity of Design for Each System

	$\bar{\Delta}$ s.u. (N)	Random	Matched & Intact Random	Nonrandom
All Δs	0.103 0.414 (341)	0.105 0.454 (117)	0.169 0.405 (132)	0.007 0.355 (92)
Audio-Tutorial	0.170 0.274 (7)	0.170 0.274 (7)		
Computer Linked	0.134 0.583 (14)	0.470 1.009 (4)	0.035 0.367 (6)	−0.053 0.114 (4)
CAI	0.010 0.743 (5)	0.143 0.941 (3)	−0.190 0.552 (2)	
CMI	0.048 0.220 (8)		0.148 0.270 (4)	−0.053 0.114 (4)
CSE	1.450 0.000 (1)	1.450 0.000 (1)		
Contracts	0.467 0.605 (12)	0.857 0.467 (7)	−0.078 0.201 (5)	
Dept. Elem. School	−0.090 0.165 (3)			−0.090 0.165 (3)
Indiv. Instr.	0.174 0.459 (131)	0.215 0.494 (56)	0.175 0.442 (53)	0.070 0.409 (22)
Mastery Learning	0.644 0.430 (13)	0.742 0.434 (10)	0.530 0.000 (1)	0.210 0.184 (2)

Table 35, continued

	$\bar{\Delta}$ s.d. (N)	Random	Matched & Intact Random	Nonrandom
All Δs	0.103 0.414 (341)	0.105 0.454 (117)	0.169 0.405 (132)	0.007 0.355 (92)
Media Based	-0.023 0.369 (100)	-0.219 0.443 (15)	0.071 0.266 (41)	-0.044 0.402 (44)
TV	0.055 0.347 (40)	0.285 0.686 (2)	0.086 0.287 (34)	-0.320 0.522 (4)
Film	-0.065 0.378 (58)	-0.283 0.407 (11)	0.000 0.107 (7)	-0.016 0.385 (40)
PSI	0.603 0.423 (15)	0.742 0.434 (10)	0.403 0.280 (3)	0.210 0.184 (2)
Prog. Instr.	0.174 0.475 (52)	0.173 0.413 (15)	0.186 0.467 (31)	0.113 0.710 (6)
Branched	0.210 0.798 (5)	0.143 0.941 (3)	0.310 0.863 (2)	
Linear	0.170 0.441 (47)	0.180 0.236 (12)	0.178 0.454 (29)	0.113 0.710 (6)
Self-Directed	0.078 0.375 (27)	0.107 0.436 (12)	0.055 0.334 (15)	
Source Papers	0.142 0.206 (13)	0.163 0.274 (3)	0.136 0.199 (10)	
Student Assisted	0.088 0.171 (6)	0.048 0.205 (4)		0.170 0.014 (2)
Team Teaching	0.058 0.378 (41)	-0.004 0.492 (14)	0.161 0.313 (19)	-0.076 0.238 (8)

Table 36
Effect Sizes by Subject Matter for Each System

	$\bar{\Delta}$ s.d. (N)	General Science	Life Science	Physical Science	Biology	Earth Science	Chemistry	Physics	Other
All Δs	0.103 0.414 (341)	0.090 0.315 (100)	0.155 0.201 (12)	0.134 0.286 (16)	0.150 0.483 (76)	0.084 0.216 (7)	0.146 0.441 (73)	−0.014 0.508 (54)	0.093 0.330 (3)
Audio-Tutorial	0.170 0.274 (7)	0.020 0.028 (2)			0.230 0.311 (5)				
Computer Linked	0.134 0.583 (14)	0.020 0.108 (3)					0.143 0.941 (3)	0.174 0.606 (8)	
CAI	0.010 0.743 (5)						0.143 0.941 (3)	−0.190 0.552 (2)	
CMI	0.048 0.220 (8)	0.020 0.108 (3)						0.064 0.279 (5)	
CSE	1.450 0.000 (1)							1.450 0.000 (1)	
Contracts	0.467 0.605 (12)				0.610 0.639 (9)		0.040 0.114 (3)		
Dept. Elem. School	−0.090 0.165 (3)	−0.090 0.165 (3)							
Indiv. Instr.	0.174 0.459 (131)	0.016 0.252 (36)	0.430 0.000 (1)	0.216 0.271 (10)	0.265 0.550 (30)	0.000 0.000 (1)	0.204 0.508 (43)	0.323 0.652 (9)	0.030 0.000 (1)
Mastery Learning	0.644 0.430 (13)				0.857 0.467 (7)		0.368 0.219 (5)	0.530 0.000 (1)	

Table 36, continued

	$\bar{\Delta}$ s.d. (N)	General Science	Life Science	Physical Science	Biology	Earth Science	Chem-istry	Physics	Other
All Δs	0.103 0.414 (341)	0.090 0.315 (100)	0.155 0.201 (12)	0.134 0.286 (16)	0.150 0.483 (76)	0.084 0.216 (7)	0.146 0.441 (73)	-0.014 0.508 (54)	0.093 0.330 (3)
Media Based	-0.023 0.369 (100)	0.066 0.328 (36)		0.096 0.159 (5)	0.149 0.477 (15)		-0.009 0.324 (15)	-0.277 0.288 (27)	0.125 0.460 (2)
TV	0.055 0.347 (40)	0.092 0.342 (26)		0.096 0.159 (5)	-0.049 0.495 (7)			-0.160 0.057 (2)	
Film	-0.065 0.378 (58)	0.090 0.244 (8)			0.323 0.414 (8)		-0.009 0.324 (15)	-0.287 0.298 (25)	0.125 0.460 (2)
PSI	0.603 0.423 (15)	0.403 0.280 (3)			0.857 0.467 (7)		0.368 0.219 (5)		
Prog. Instr.	0.174 0.475 (52)	-0.065 0.342 (10)	0.430 0.000 (1)	0.148 0.161 (4)	0.055 0.424 (12)		0.291 0.550 (22)	0.533 0.516 (3)	
Branched	0.210 0.798 (5)						0.210 0.798 (5)		
Linear	0.170 0.441 (47)	-0.065 0.342 (10)	0.430 0.000 (1)	0.148 0.161 (4)	0.055 0.424 (12)		0.315 0.485 (17)	0.533 0.516 (3)	
Self-Directed	0.078 0.375 (27)	0.200 0.263 (7)			0.172 0.479 (6)	0.000 0.000 (1)	-0.047 0.355 (10)	0.050 0.570 (3)	
Source Papers	0.142 0.206 (13)	0.163 0.274 (3)	0.136 0.199 (10)						
Student Assisted	0.088 0.171 (6)	0.088 0.171 (6)							
Team Teaching	0.058 0.378 (41)	0.295 0.389 (4)	0.136 0.199 (10)	-0.490 0.000 (1)	0.062 0.487 (16)		-0.027 0.286 (3)	-0.081 0.270 (7)	

Table 37
Effect Sizes by Immediate or Retention Measures for Each System

	$\bar{\Delta}$ s.d. (N)	IMMEDIATE	RETENTION
All Δs	0.103 0.414 (341)	0.126 0.430 (290)	-0.093 0.250 (33)
Audio-Tutorial	0.170 0.274 (7)	0.223 0.268 (3)	(0)
Computer Linked	0.134 0.583 (14)	0.221 0.626 (11)	-0.183 0.240 (3)
CAI	0.010 0.743 (5)	0.118 0.812 (4)	-0.420 0.000 (1)
CMI	0.048 0.220 (8)	0.085 0.234 (6)	-0.065 0.177 (2)
CSE	1.450 0.000 (1)	1.450 0.000 (1)	(0)
Contracts	0.467 0.605 (12)	0.522 0.603 (11)	-0.130 0.000 (1)
Dept. Elem. School	-0.090 0.165 (3)	-0.090 0.165 (3)	(0)
Indiv. Instr.	0.174 0.459 (131)	0.220 0.482 (108)	-0.109 0.234 (12)
Mastery Learning	0.644 0.430 (13)	0.644 0.430 (13)	(0)
Media Based	-0.023 0.369 (100)	-0.009 0.377 (85)	-0.133 0.347 (12)
TV	0.055 0.347 (40)	0.055 0.347 (40)	(0)
Film	-0.065 0.378 (58)	-0.051 0.399 (43)	-0.133 0.347 (12)
PSI	0.603 0.423 (15)	0.603 0.423 (15)	(0)
Prog. Instr.	0.174 0.475 (52)	0.260 0.497 (40)	-0.113 0.276 (8)
Branched	0.210 0.798 (5)	0.590 0.854 (3)	-0.360 0.085 (2)
Linear	0.170 0.441 (47)	0.234 0.467 (37)	-0.030 0.269 (6)
Self-Directed	0.078 0.375 (27)	0.095 0.396 (20)	-0.050 0.523 (2)
Source Papers	0.142 0.206 (13)	0.142 0.206 (13)	(0)
Student Assisted	0.088 0.171 (6)	0.088 0.171 (6)	(0)
Team Teaching	0.058 0.378 (41)	0.063 0.398 (37)	0.035 0.007 (2)

Table 38
Effect Sizes by Type of Outcome Criterion
for Each System

	$\bar{\Delta}$ s.d. (N)	Cognitive	Affective	Science Methods	Psycho-Motor	Critical Thinking	Creativity	Self Concept	Logical Thinking
All Δs	0.103 0.414 (341)	0.069 0.407 (249)	0.034 0.310 (45)	0.299 0.415 (19)	0.892 0.684 (6)	0.234 0.311 (7)	0.430 0.457 (4)	0.317 0.100 (3)	0.403 0.280 (3)
Audio-Tutorial	0.170 0.274 (7)	0.088 0.287 (5)	0.330 0.000 (1)					0.420 0.000 (1)	
Computer Linked	0.134 0.583 (14)	0.216 0.618 (11)	-0.167 0.359 (3)						
CAI	0.010 0.743 (5)	0.158 0.769 (4)	-0.580 0.000 (1)						
CMI	0.048 0.220 (8)	0.050 0.260 (6)	0.040 0.028 (2)						
CSE	1.450 0.000 (1)	1.450 0.000 (1)							
Contracts	0.467 0.605 (12)	0.218 0.569 (5)	0.330 0.449 (3)	1.235 0.714 (2)		0.530 0.509 (2)			
Dept. Elem. School	-0.090 0.165 (3)	-0.090 0.165 (3)							
Indiv. Instr.	0.174 0.459 (131)	0.118 0.440 (102)	0.160 0.373 (10)	0.428 0.565 (9)	1.165 0.064 (2)	0.325 0.405 (4)	0.495 0.530 (2)	0.365 0.078 (2)	
Mastery Learning	0.644 0.430 (13)	0.498 0.278 (8)	0.515 0.446 (2)	1.235 0.714 (2)		0.890 0.000 (1)			

Table 38, continued

	$\bar{\Delta}$ s.d. (N)	Cognitive	Affective	Science Methods	Psycho-motor	Critical Thinking	Crea-tivity	Self Concept	Logical Thinking
All Δs	0.103 0.414 (341)	0.069 0.407 (249)	0.034 0.310 (45)	0.299 0.415 (19)	0.892 0.684 (6)	0.234 0.311 (7)	0.430 0.457 (4)	0.317 0.100 (3)	0.403 0.280 (3)
Media Based	-0.023 0.369 (100)	-0.030 0.388 (75)	-0.104 0.298 (16)	0.118 0.143 (5)	-0.080 0.000 (1)	0.160 0.014 (2)	0.770 0.000 (1)		
TV	0.055 0.347 (40)	0.022 0.355 (33)	-0.120 0.000 (1)	0.173 0.087 (4)		0.150 0.000 (1)	0.770 0.000 (1)		
Film	-0.065 0.378 (58)	-0.055 0.416 (40)	-0.103 0.309 (15)	-0.100 0.000 (1)	-0.080 0.000 (1)	0.170 0.000 (1)			
PSI	0.603 0.423 (15)	0.493 0.300 (7)	0.515 0.446 (2)	1.235 0.714 (2)		0.890 0.000 (1)			0.403 0.280 (3)
Prog. Instr.	0.174 0.475 (52)	0.173 0.479 (51)	0.200 0.000 (1)						
Branched	0.210 0.798 (5)	0.210 0.798 (5)							
Linear	0.170 0.441 (47)	0.169 0.446 (46)	0.200 0.000 (1)						
Self-Directed	0.078 0.375 (27)	-0.018 0.341 (16)	-0.097 0.458 (3)	-0.110 0.000 (1)		0.170 0.000 (1)	0.495 0.530 (2)	0.420 0.000 (1)	0.403 0.280 (3)
Source Papers	0.142 0.206 (13)	0.142 0.171 (9)	-0.190 0.000 (1)	0.253 0.253 (3)					
Student Assisted	0.088 0.171 (6)	0.105 0.332 (2)	0.170 0.014 (2)			0.020 0.000 (1)	-0.040 0.000 (1)		
Team Teaching	0.058 0.378 (41)	0.087 0.400 (31)	-0.124 0.235 (7)	0.183 0.177 (3)					

Table 39
Effect Sizes by Method of Measurement for Each System

	$\bar{\Delta}$ s.d. (N)	Published	Modified Publ. & Ad hoc	Other Assessment
All Δs	0.103 0.414 (341)	0.045 0.387 (173)	0.126 0.393 (158)	0.951 0.486 (8)
Audio-Tutorial	0.170 0.274 (7)	0.375 0.064 (2)	0.088 0.287 (5)	
Computer Linked	0.134 0.583 (14)	-0.158 0.256 (5)	0.297 0.661 (9)	
CAI	0.010 0.743 (5)	-0.580 0.000 (1)	0.158 0.769 (4)	
CMI	0.048 0.220 (8)	-0.053 0.114 (4)	0.148 0.270 (4)	
CSE	1.450 0.000 (1)		1.450 0.000 (1)	
Contracts	0.467 0.605 (12)	0.467 0.605 (12)		
Dept. Elem. School	-0.090 0.165 (3)	-0.090 0.165 (3)		
Indiv. Instr.	0.174 0.459 (131)	0.159 0.442 (65)	0.159 0.453 (64)	1.165 0.064 (2)
Mastery Learning	0.644 0.430 (13)	0.713 0.500 (9)	0.488 0.161 (4)	

Table 39, continued

	$\bar{\Delta}$ s.d. (N)	Published	Modified Publ. & Ad hoc	Other Assessment
All Δs	0.103 0.414 (341)	0.045 0.387 (173)	0.126 0.393 (158)	0.951 0.486 (8)
Media Based	-0.023 0.369 (100)	-0.081 0.351 (51)	0.038 0.381 (49)	
TV	0.055 0.347 (40)	0.020 0.119 (7)	0.063 0.379 (33)	
Film	-0.065 0.378 (58)	-0.084 0.377 (42)	-0.014 0.390 (16)	
PSI	0.603 0.423 (15)	0.713 0.500 (9)	0.438 0.219 (6)	
Prog. Instr.	0.174 0.475 (52)	0.258 0.394 (10)	0.154 0.494 (42)	
Branched	0.210 0.798 (5)		0.210 0.798 (5)	
Linear	0.170 0.441 (47)	0.258 0.394 (10)	0.146 0.455 (37)	
Self-Directed	0.078 0.375 (27)	0.088 0.392 (16)	0.065 0.368 (11)	
Source Papers	0.142 0.206 (13)	0.183 0.220 (10)	0.007 0.006 (3)	
Student Assisted	0.088 0.171 (6)	-0.050 0.076 (3)	0.227 0.099 (3)	
Team Teaching	0.058 0.378 (41)	0.094 0.394 (23)	0.014 0.361 (16)	

Table 40
Effect Sizes by Calculation of Effect Size for Each System

	$\bar{\Delta}$ s.d. (N)	From raw data	By direct calculation	Less trustworthy methods
All Δs	0.103 0.414 (341)	0.099 0.435 (179)	0.144 0.422 (117)	0.013 0.275 (45)
Audio-Tutorial	0.170 0.274 (7)	0.130 0.312 (4)	0.335 0.262 (2)	0.000 0.000 (1)
Computer Linked	0.134 0.583 (14)	0.149 0.640 (11)	-0.145 0.064 (2)	0.530 0.000 (1)
CAI	0.010 0.743 (5)	0.010 0.743 (5)		
CMI	0.048 0.220 (8)	0.028 0.079 (5)	-0.145 0.064 (2)	0.530 0.000 (1)
CSE	1.450 0.000 (1)	1.450 0.000 (1)		
Contracts	0.467 0.605 (12)	-0.078 0.201 (5)	0.857 0.467 (7)	
Dept. Elem. School	-0.090 0.165 (3)	0.080 0.000 (1)	-0.250 0.000 (1)	-0.100 0.000 (1)
Indiv. Instr.	0.174 0.459 (131)	0.176 0.476 (72)	0.236 0.469 (45)	-0.032 0.258 (14)
Mastery Learning	0.644 0.430 (13)	0.473 0.194 (3)	0.857 0.467 (7)	0.317 0.226 (3)

Table 40, continued

	$\bar{\Delta}$ s.d. (N)	From raw data	By direct calculation	Less trust-worthy methods
All Δs	0.103	0.099	0.144	0.013
	0.414	0.435	0.422	0.275
	(341)	(179)	(117)	(45)
Media Based	-0.023	-0.080	0.055	-0.097
	0.369	0.345	0.413	0.244
	(100)	(42)	(44)	(14)
TV	0.055	0.018	0.066	0.096
	0.347	0.428	0.341	0.159
	(40)	(12)	(23)	(5)
Film	-0.065	-0.101	0.044	-0.204
	0.378	0.308	0.488	0.219
	(58)	(28)	(21)	(9)
PSI	0.603	0.438	0.857	0.210
	0.423	0.219	0.467	0.184
	(15)	(6)	(7)	(2)
Prog. Instr.	0.174	0.173	0.373	-0.207
	0.475	0.485	0.420	0.140
	(52)	(43)	(6)	(3)
Branched	0.210	0.210		
	0.798	0.798		
	(5)	(5)		
Linear	0.170	0.168	0.373	-0.207
	0.441	0.445	0.420	0.140
	(47)	(38)	(6)	(3)
Self-Directed	0.078	0.079	0.495	-0.092
	0.375	0.348	0.530	0.386
	(27)	(20)	(2)	(5)
Source Papers	0.142		0.163	0.136
	0.206		0.274	0.199
	(13)		(3)	(10)
Student Assisted	0.088		0.088	
	0.171		0.171	
	(6)		(6)	
Team Teaching	0.058	-0.101	0.253	0.122
	0.378	0.374	0.479	0.239
	(41)	(17)	(9)	(15)

Table 41
Effect Sizes by Source of Means for Each System

	$\bar{\Delta}$ s.d. (N)	Unadjusted Posttest	Covariance Adjusted	Pre post Difference	Other
All Δs	0.103 0.414 (341)	0.125 0.448 (162)	0.086 0.387 (67)	0.087 0.382 (93)	0.024 0.358 (18)
Audio-Tutorial	0.170 0.274 (7)	0.0400 0.0000 (1)		0.230 0.311 (5)	0.000 0.000 (1)
Computer Linked	0.134 0.583 (14)	-0.174 0.269 (8)		0.548 0.731 (5)	0.530 0.000 (1)
CAI	0.010 0.743 (5)	-0.295 0.341 (4)		1.230 0.000 (1)	
CMI	0.048 0.220 (8)	-0.053 0.114 (4)		0.020 0.108 (3)	0.530 0.000 (1)
CSE	1.450 0.000 (1)			1.450 0.000 (1)	
Contracts	0.467 0.605 (12)	-0.078 0.201 (5)	0.857 0.467 (7)		
Dept. Elem. School	-0.090 0.165 (3)		-0.010 0.127 (2)	-0.250 0.000 (1)	
Indiv. Instr.	0.174 0.459 (131)	0.176 0.514 (51)	0.198 0.467 (32)	0.150 0.412 (40)	0.190 0.330 (8)
Mastery Learning	0.644 0.430 (13)	0.368 0.219 (5)	0.857 0.467 (7)		0.530 0.000 (1)

Table 41, continued

	$\bar{\Delta}$ s.d. (N)	Unadjusted Posttest	Covariance Adjusted	Pre-post Difference	Other
All Δs	0.103 0.414 (341)	0.125 0.448 (162)	0.086 0.387 (67)	0.087 0.382 (93)	0.024 0.358 (18)
Media Based	-0.023 0.369 (100)	-0.042 0.393 (40)	-0.048 0.279 (26)	0.071 0.416 (28)	-0.225 0.250 (6)
TV	0.055 0.347 (40)	0.018 0.428 (12)	0.144 0.171 (9)	0.041 0.372 (18)	-0.040 0.000 (1)
Film	-0.065 0.378 (58)	-0.044 0.386 (26)	-0.149 0.274 (17)	0.125 0.503 (10)	-0.262 0.260 (5)
PSI	0.603 0.423 (15)	0.381 0.224 (8)	0.857 0.467 (7)		
Prog. Instr.	0.174 0.475 (52)	0.242 0.495 (30)	-0.003 0.477 (3)	0.095 0.446 (19)	
Branched	0.210 0.798 (5)	-0.400 0.028 (2)		0.617 0.809 (3)	
Linear	0.170 0.441 (47)	0.288 0.480 (28)	-0.003 0.477 (3)	-0.003 0.295 (16)	
Self-Directed	0.078 0.375 (27)	0.063 0.374 (15)	-0.275 0.106 (2)	0.027 0.315 (7)	0.507 0.376 (3)
Source Papers	0.142 0.206 (13)			0.142 0.206 (13)	
Student Assisted	0.088 0.171 (6)	0.048 0.205 (4)		0.170 0.014 (2)	
Team Teaching	0.058 0.583 (41)	0.083 0.507 (17)	0.060 0.095 (3)	0.021 0.270 (16)	0.094 0.330 (5)

CONCLUSIONS

Although it must be done with caution, it is possible to draw some broad generalizations from the integration of research studies on science instructional systems. The most successful innovative systems appear to be mastery learning ($\bar{\Delta} = .64$ overall and $\bar{\Delta} = .50$ for cognitive achievement) and P.S.I. ($\bar{\Delta} = .60$ overall and $\bar{\Delta}$ - .49 for cognitive achievement). Specific data on the various other outcome variables displayed in Table 6 verify that, in addition to being approximately one-half standard deviation better than control groups on cognitive measures, these two systems look good on other variable as well. On the other hand, media based systems in general appear to perform at a lower level than the traditional instruction used as the control group treatment. Most of the remaining systems operate at a level very little higher than the conventional instructions they have replaced; most have an average effect size approximating 0.1 standard deviations both on outcome measures overall and on cognitive measures. When compared with conventional instruction, instructional systems do not show a striking advantage ($\bar{\Delta} = .10$) and the impact in terms of affective measures is practically nothing ($\bar{\Delta} = .04$).

Making additional broad generalizations is difficult because of the small number of effect sizes found for many outcome variables. In addition, the number of different instructional systems for which there is data on a given outcome variable is generally very small. As a result, it is difficult to make generalizations about instructional systems broadly, since the data provided in this meta-analysis is limited to only a few instructional systems for a particular outcome variable. In the case of three variables, however, some generalizations may be possible.

For science methods, critical thinking, and logical thinking, the number of effect sizes is large enough, and the diversity of teaching systems evaluated with respect to a particular outcome variable is diverse enough, at one can say something about instructional systems in general. A review of Table 10b indicates there is an average effect size in favor of the instructional system of approximately .40 for these three outcome variables.

The most important conclusions of this meta-analysis however, do not pertain to instructional systems overall but to particular systems. The data in Table 10b is instructive in this regard. A potential recording of data for an outcome variable under more than one instructional system in Table 10b is not a concern because data (whether used for another instructional system also or not) are indicative of the impact of the particular instructional system under consideration.

A related point is that meaningful interpretation of the results of this meta-analysis with respect to a given instructional system requires careful analysis and examination of that system. One must know what it is about each system that makes it work, and in particular what it is that the most successful systems have in common. Such a review requires that one look at the characteristics of the various systems and determine what makes each one successful enough to stand out. An example of such an endeavor is meta-analysis work done in higher education (Kulik & Kulik, 1979) which identifed P.S.I. and some other instructional systems as being useful on the college level. In their examination of these instructional systems, they stated that a key characteristic held in common by these successful approaches was frequent testing with immediate feedback. While it is pleasing to see commonality between the results of the meta-analysis reported here and the work of another researcher at the college level, the key point to be made here is that the interpreter of these results must look beyond simple labels or even rather extended definitions as reported in this paper and analyze carefully what the components of each instructional system are. Such careful analysis work may make it possible to identify the key facets of instructional systems which are essential for their success. The results of such interpretive work are of value to practitioners in the field and to researchers needing to identify the elements of instruction with the st potential for increasing learning.

SECTION IV. STUDIES INCLUDED IN THIS REPORT

NUMERICAL LIST OF CODED STUDIES

(For more details, see alphabetical list by authors)

Number	Author	Source*	Measure(s) Used
2001	Anderson, C. J.	4	CHEM Study Chemistry, Chapters 1-7
2002	Williams, W. W.	4	Comprehensive Test in Basic Physical Science
2003	Charles, E.	4	Ad hoc** Cloze tests
2004	Young, P. A.	4	Ad hoc Biology achievement
2005	Koenig, H. G.	4	Modified Minnesota High School Achievement Exam
2006	Grooms, H. H.	4	Metropolitan Achievement Test
2007	Wachs, S. R.	4	Test on Understanding Science (TOUS), Form Jx
			Ad hoc physics exam
			Ad hoc biology exam
2008	Ward, P. E.	4	STEP Science, Form A
2009	Williams, H. R.	4	ACS-NSTA Cooperative Exam, High School Chemistry, Form 1961
			TOUS
			Thurstone Interest Schedule
			Purdue Master Attitude Scale
2010	Krockover, G. H.	4	ACS Coop. Exam, General Chemistry, Form 1963
			TOUS, Form 2
			Watson-Glaser Test of Critical Thinking
			Prouse Subject Preference Survey

---

*1=journal; 2=book; 3=master's thesis; 4=doctoral dissertation; 5=unpublished; 6=paper presented at a conference.

**"Ad hoc" indicates instruments created by the investigator for the study.

2011	Scarpino, F. L.	4	Co-Op Chemistry Test, Forms A, B.
			Anderson-Fisk Chemistry Test, Forms E, F.
			DuBelle Student Preference Report, Forms A, B.
			Ad hoc lab skills test
2012	Moore, B. F.	4	Brown-Holtzman Survey of Study Habits & Attitudes
			BSCS Comprehensive Final Exam
2013	Fryar, W. R.	4	Ad hoc, "Aquatic Life"
2014	Eshleman, W. H.	4	Ad hoc cognitive test
2015	Marshall, G.	4	Ad hoc cognitive test
			School midterm exams
			School final exam
2016	Darnowski, V. S.	4	Ad hoc cognitive test
2017	McKee, R. J.	4	Ad hoc cognitive test
			Ad hoc application test
2018	Joslin, P. H.	4	N.Y. State Regent's Exam
			Ad hoc cognitive test
2019	Dasenbrock, D. H.	4	Ad hoc cognitive test
2020	Molotsky, L. L.	4	Ad hoc Biol. Achievement Test
2021	Inventash, H.	4	Co-Op Science Test
			STEP Achievement Tests, Forms 3B, 3A.
			TOUS
2022	Humphreys, D. W.	4	BSCS Comprehensive Exam
			Ad hoc Q-sort
2023	Stedman, C. H.	4	Ad hoc general achievement test

2024	Turpin, G. R.	4	Ad hoc semantic differential scale
2025	Thornton, W. T.	4	Ad hoc achievement test
2026	Waine, S. I.	4	Chemistry I: Atomic Structure and Bonding
2027	Meiller, R. D.	4	Nelson Biology Test
			Purdue Student Attitude Test
			Dunning Physics Test
2028	Summerlin, L. R.	4	Ad hoc achievement test
2029	Wiegand, C. H.	4	Every Pupil Achievement Test, Elementary Science, Grades V-VIII.
2030	James, R. K.	4	Ad hoc Seventh Grade Matter Final Exam
			Metropolitan Achievement Test--Adv. Science Test
			Anderson-Fisk Chemistry Test, Form E.
			Read General Science Test
			Facts About Science Test
			Watson-Glaser Critical Thinking Appraisal
			TOUS
2031	Slattery, J. B.	4	New York Regents Exam, Biology
			New York Regents Exam, Chem.
			New York Regents Exam, Physics
2032	Braly, J. L.	4	ACS-NSTA Chemistry Exam
2033	Reed, L. H.	4	Stanford Achievement Test: Science.
			Remmer's Attitude Toward Any School Subject Scale
			Piers-Harris Children's Self-Concept Scale

2034	Koch, D. P.	4	Project Physics Achievement Test, Units 2, 3.
			Ad hoc confidence scores
2035	Taffel, A.	4	Ad Hoc Midyear Achievement Test
			New York Regents Exam, Physics
			Dunning Physics Test
2036	Call, R. L.	4	ACS-NSTA Cooperative Exam, High School Chemistry
			Anderson-Fisk Chemistry Test, Form F
2037	Blank, S. S.	4	Ad hoc cognitive test
2038	Mottillo, J. L.	4	Ad hoc unit exams
2039	Denton, J. J.	4	Ad hoc Physics Achievement Test
			Purdue Master Attitude Scale for Measuring Attitude Toward Any School Subject, Form B
2040	Payne, C. R.	4	Ad hoc chapter tests
2041	White, R. W.	4	Cooperative Biology
2042	Heffernan, D. F.	4	TOUS
			Watson-Glaser Critical Thinking Appraisal
2043	Carnes, P. E.	4	Ad hoc cognitive tests
2044	O'Toole, R. J.	4	Ad hoc problem solving test
			Ad hoc cognitive tests
2045	Connor, J. L.	4	Ad hoc achievement test
2046	Breedlove, C. B.	4	TOUS
			Allen Attitude Inventory
			Ad hoc achievement test
2047	Hunt, E. G.	4	California Survey Test in Physical Science

2048	Alcorta, L. B.	4	Iowa Test of Educational Development #6, #2 Brown-Holtzmann Survey of Study Habits STEP, Forms 3A, 3B, 2A, 2B Nelson Biology Test Anderson Chemistry Test Ad hoc achievement test Facts About Science Test, A and B
2049	Aaron, G.	4	Ad hoc cognitive test
2050	Love, G. H.	4	Stanford Science Achievement Test, Form X
2051	Brown, F. K., & D. P. Butts	1	Ad hoc cognitive test
2052	Raghubir, K. P.	1	Ad hoc achievement test
2053	Long, J. C., J. R. Okey, & R. H. Yeany	1	Ad hoc achievement test
2054	Martin, W. J., & P. E. Bell	1	Ad hoc evaluations, lab skills and affective
2055	Study deleted		
2056	Toohey, J. V.	1	Ad hoc achievement test
2057	Welliver, P. W.	1	Ad hoc Physical Science Achievement Test TOUS Ad hoc Science Current Events Test STEP Science, Form 3A Thurstone Interest Scale
2058	Gallagher, J. J.	1	Coded videotapes Ad hoc test of interaction recognition

2059	Anderson, C., & D. Butts	1	Ad hoc achievement test
			Ad hoc Attitude to Science questionnaire
2060	Anderson, R. D., & A. R. Thompson	1	Ad hoc achievement test
			Stanford Achievement Test
			Boulder Test of Creative Thinking
			Boulder Test of Critical Thinking
2061	Netburn, A. N.	4	Ad hoc cognitive test
			Time spent on each lesson
2062	Cowan, P. J.	4	PSSC tests, 1-5
2063	Siddiqi, M. N.	4	PSSC Tests, 1-5
2064	Galey, M.	4	Ad hoc achievement test, performance interview
2065	Tucker, J. L.	4	Ad hoc Picture Test for Science Processes
			Ad hoc Science Concepts Test
2066	Fulton, H. F.	1	BSCS Final Exam, 1964
			Nelson Biology Test, Form E
			TOUS, Form W
			F.A.S.
			Watson-Glaser Critical Thinking Appraisal, Form Zm
			Silance Attitude Scale, Form A
			Prouse Subject Preference Survey
2067	Wash, J. A.	4	Ad hoc achievement test
2068	Pella, M. O., & C. Poulos	1	Ad hoc Biology Exam
			Coop. Biology Exam

2069	Sutman, F. X., & M. Yost	1	Ad hoc unit tests
2070	Hedges, W. D., & M. A. MacDougall	1	STEP Science Achievement, Forms A and B California Interest Inventory
2071	Garside, L. J.	4	Ohio Physics Test Wisconsin Physics Final Test Physics Accumulated Test
2072	Anderson, K. E., F. S. Montgomery, & R. W. Ridgway	1	Minnesota State Board Exam in Biology, 1947
2073	Anderson, K. E., F. S. Montgomery, & H. A. Smith, & D. S. Anderson	1	Nelson Biology Test, Forms Am, Bm
2074	Anderson, K. E., & F. S.	1	Dunning Physics Test, Forms Am and Bm
2075	Jacobs, H. N., & J. K. Bollenbacher	1	Coop. Biology Test, Forms X, Y Test of Knowledge About Science and Scientists
2076	Jacobs, L. C.	4	Ad hoc achievement tests
2077	Strehle, J. A.	4	Ad hoc achievement test
2078	Walker, M. A.	4	Otis-Lennon Mental Ability Test, Elem. II Level, Form K
2079	Przekop, L. R.	4	Ad hoc achievement test
2080	Popham, W. J., & J. M. Sadnavitch	1	Coop. Physics Test, 1950 Coop. Chemistry Test Thurstone Interest Schedule, 1947 Scale for Measuring Attitude Toward Any School Subject
	Also included in #2080: Sadnavitch, Popham, & Black	1	Coop. Physics Test Coop. Chemistry Test

2081	Pella, M. O., J. Stanley, C. A. Wedemeyer, & W. A. Wittich	1	Wisconsin Physics Test
			Ohio Physics Schol. Test
			Ad hoc affective
2082	Allison, R. W.	4	Adaptation of Allen Inventory of Attitudes Toward Science and Scientific Careers
2083	Dilorenzo, L. T., & J. W. Halliwell	1	Metropolitan Achievement Test
2084	Crabtree, J. F.	1	Ad hoc achievement test
			Time on task
2085	Hug, W. E.	1	Comprehensive Final Exam. in First Year Biology
2086	O'Brien, S. J.	4	Childhood Attitude Toward Problem Solving
2087	Troost, C. J., & S. Morris	1	Ad hoc cognitive measure
2088	Beets, M. M.	4	Creative Thinking Test (adapted for elementary)
2089	Beisenherz, P. C.	4	Picture Test for Science Processes, Grades 1 & 2; 3&4 (local)
			Science Concepts Test, Grades 3 & 4 (local)
2090	Champa, V. A.	4	Coop. Science Test
2091	Grassell, E. M.	4	STEP Physics Test
			Dunning Physics Test
			Ad hoc achievement tests
2092	Boblick, J. M.	1	Ad hoc achievement test
			Time on task
2093	Boblick, J. M.	1	Ad hoc achievement test
2094	Wickline, L. E.	4	Allen Attitude Scale
			Facts About Science Test

2095	Nordland, F. H., J. B, Kahle, S. Randak, & T. Watts	1	Ad hoc unit tests
2096	Penick, J. E., D. Schlitt, S. Bender, & J. Lewis	1	Torrance Test of Creative Thinking (Figural Creativity; Verbal Creativity)
2097	Kahle, J. B., F. H. Nordland, and C. B. Douglass	1	Ad hoc achievement test
2098	Penn, R. F.	4	ACS/CHEM test
			TOUS
			Cornell Critical Thinking Test, Form Z (1961)
2099	Johnson, L.	4	Ad hoc achievement test
2100	May, J.	1	ACS-NSTA Test, form 1971
			ACS-NSTA Test, Form 1970 advanced
2101	Richard, P. W.	1	BSCS Achievement Test
2102	Kline, A. A.	1	Ad hoc achievement test
2103	Fiel, R. L., & J. R. Okey	1	Ad hoc achievement test
2104	Zeschke, R.	1	Ad hoc achievement test
			Time on task
2105	Crocker, R. K., et al.	5	Ad hoc achievement tests
2106	Black, W. A., et al.	5	Cooperative Chemistry Test
			Cooperative Phyics Test
2107	Yarber, W. L.	1	A Venereal Disease Knowledge Inventory
2108	Patterson, M. D.	6	Bristol Study Skills (abbreviated)
2109	Monaco, W. J., & M. Szabo	6	Stanford Achievement Test, Science Sub-score

2110	Linn, M. C., B. Chen, & H. D. Thier	5	Science Process Test--Variables; Experimentation Interviews
2111	Fritz, J. O.	5	Ad hoc achievement test Adapted Allen Attitude Scale Toward Science and Scientific Careers
2112	Garry, R. J., H. J. Dietmeyer, M. Kraft, & A. C. Sheehan	5	Ad hoc science information test Science Reasoning test (local) STEP Science Test
2113	Winter, S. S., S. D. Farr, J. J. Montean, & J. A. Schmidt	5	New York Regents Exam, Chemistry Science Reasoning Test Kuder Preference Inventory
2114	Vandermeer, A. W.	5	Calvert Science Information Test, Intermediate Form B Ad hoc unit tests
2115	Robinson, D. B.	5	New York Regents Exam, Biology, 1967 Nelson Biology Test
2116	Denton, J. J., & F. J. Gies	6	Number of objectives achieved
2117	Swanson, D. H.	6	Adaptation of previous Regents Exams, Chemistry Number of objectives mastered
2118	Strevell, W. H.	5	Dunning Physics Achievement Test
2119	Nelson, C. M.	1	Ad hoc achievement test
2120	Wade, S. E.	1	Ad hoc achievement tests
2121	McCollum, T. E.,	1	Achievement tests

2122	Noall, M. F., & L. Winget	1	Coop. Physics Test, Forms X, Y Strong Vocational Interest Blank (Attitude to Science)
2123	Glass, L. W., & R. E. Yager	1	TOUS FAS
2124	Simmons, J. B., W. J. Davis, G. C. Ramseyer, & J. J. Johnson	1	BSCS Standardized Biology Achievement Test
2125	Anderson, K. E., F. S. Montgomery, & S. F. Moore	1	Anderson Chemistry Test, Form Am ACS-NSTA Chemistry Exam, Form 1959 Ad hoc lab skills test
2126	Jerkins, K. F.	4	TOUS Coop. Science Test MPATI
2127	Lee, J. E.	4	Ad hoc biology achievement test Attitude Toward Subject Self Concept of Ability in Science
2128	Shinfeld, S. L.	4	ACS Chemistry Test Ad hoc achievement tests
2129	Moore, W. J.	4	Silance: Attitude Toward Any School Subject Processes of Science Test Final Comprehensive Exam (BSCS Patterns & Processes)
2130	Martinez-Perez, L.	4	Piers-Harris Children's Self-Concept Scale Attitude Toward Science Inventory (modified)
2131	Hughes, W. R.	4	Ad hoc Processes of Science Test; Content Examinations.

REFERENCES

Block, J. H. Mastery learning theory and practice. New York: Holt, Rinehart & Winston, 1971.

Carmichael, J. W., Jr. General chemistry by PSI at a Minority Institution. Journal of Chemical Education, 1976, 53(12):791-2.

Emerson, David W. Teaching organic chemistry by a modified Keller plan. Journal of Chemical Education, 1975, 52(4):228-9.

Good, Carter V. (Ed.). Dictionary of education. New York: McGraw-Hill, 1973.

Hartley, J. R. Computer assisted learning in the sciences: some progress and some prospects. Studies in Science Education, 3(1976):69-96.

Herring, J. Dudley, Harold H. Jaus, Van Neie, Thom Luce, & Terry O'Heron. A summary of research in science teaching--1974. Purdue University. (ERIC Document Reproduction Service No. ED 116 965)

Jackson, Gregg B. Methods for reviewing and integrating research in the social sciences. The George Washington University, Social Research Group, 1978.

Kulik, James A. and Kulik, Chen-Lin C. "College Teaching" in Peterson, P.L. and Walberg, H.J. (ed.) Research on Teaching - Berkeley, California: McCutchon Publishers, 1979.

Kuska, Henry A. Use of the Keller plan in a freshman chemistry trailer course. Journal of Chemical Education, 1976, 53(8):505.

Marchese, Richard W. A literature search and review of the comparison of individualized and conventional modes of instruction in science. School Science and Mathematics, 1977, 77(8):699-703.

Melton, Raymond G. Summary of the findings of three NSF sponsored studies on the status of precollege science education. In National Science Foundation, What are the needs in precollege science, mathematics, and social science education? Views from the field. Washington: Government Printing Office, 1980.

Nordland, Floyd, Jane Kahla, & Hugh Via. A practical approach to an audio-tutorial system. School Science and Mathematics, 1972, 72(8):673-678.

Novak, Joseph D. A summary of research in science education for 1972. Ithaca, NY: Cornell University, 1974. (ERIC Document Reproduction Service No. ED 090 055)

Palladino, George F. General chemistry: an alternative to PSI for advanced students. Journal of Chemical Education, 1979, 56(5):323-324.

Postlethwaite, S. N., J. Novak, & H. Murray. An integrated experience approach to learning. Purdue University, Lafayette, Indiana. Minneapolis: Burgess, 1964. (ERIC Document Reproduction Service No. ED 010 765)

Ramsey, Gregor A., & Robert W. Howe. An analysis of research on instructional procedures in secondary school science, part II, instructional procedures. Science Teacher, 1969, 36(4):72-91.

Schramm, Wilbur. Programed instruction today and tomorrow. Reprinted in Four Case Studies of Programed Instruction. New York: Fund for the Advancement of Education, 1964.

Silberman, Harry F. Characteristics of some recent studies of instructional methods. In John E. Coulson (Ed.), Programmed Learning and Computer-Based Instruction. New York: John Wiley and Sons, 1962.

Smith, Homer A., Jr. The evolution of a self-paced organic chemistry course. Journal of Chemical Education, 1976, 53(8):510-512.

Weiss, Iris R. 1977 national survey of science, mathematics and social science education highlights report. In The status of pre-college science, mathematics, and social studies educational practices in U.S. schools: An overview and summaries of three studies. Washington: National Science Foundation, 1978.

THE EFFECTS OF VARIOUS SCIENCE TEACHING STRATEGIES ON ACHIEVEMENT

Kevin C. Wise

James R. Okey

University of Georgia
Athens, GA 30602

TABLE OF CONTENTS

LIST OF TABLES

THE EFFECTS OF VARIOUS SCIENCE TEACHING ST..ATEGIES ON ACHIEVEMENT

Kevin C. Wise and James R. Okey
University of Georgia
Athens, GA 30602

Purpose of the Study

The purpose of this study was to synthesize findings of the effects on science achievement of various teaching strategies using the procedures of meta-analysis (Glass, 1976, 1978). This is one of seven areas of science education research selected for study in the University of Colorado Science Meta-Analysis Project, The seven areas had been selected by the Colorado project, in consultation with a national panel, as representing significant blocks of findings of sufficient importance and scope to justify an integration of the literature.

Numerous studies dealing with the effects of various science teaching techniques exist in a variety of documents. The integration and interpretation of this aggregate research on the topic of science teaching techniques cannot be handled sufficiently through narrative means alone. Chronologically ordered verbal descriptions may be sufficient when only a few studies are related to a topic. When tens or hundreds of studies are involved, however, a narrative approach fails to accommodate the accumulated knowledge.

If the findings of many studies are regarded as data points and addressed in a statistical manner, they can be integrated using meta-analysis. Through this technique information can be compiled from many studies that when taken as a group in the narrative sense appear inconclusive and even incomprehensible.

The techniques of meta-analysis have been extensively described elsewhere (Glass, 1978; Glass, McGaw, White, & Smith, 1980). Integration of research findings have been done in areas such as class size and achievement (Glass & Smith, 1979), diagnostic remedial instruction in science (Yeany & Miller, 1981), and attitude and achievement in science (Willson, 1981). In its most basic form the procedure of meta-analysis involves determining the difference between experimental and control group mean scores in standard deviation units (called an effect size). The impact of a technique (such as a particular teaching strategy) in standard score units can then be examined across a variety of studies.

In this meta-analysis study, the impact of twelve categories of teaching techniques were examined. In addition to calculating the size of the effect in each study, information was collected about student and teacher characteristics, details of the treatment, and experimental conditions. The purpose of collecting this contextual information was to determine the circumstances in which the teaching strategies had their influence. By crosstabulating various features of studies (e.g., size of class or grade level of subjects) with the effect size a picture of the conditions under which a teaching strategy has maximum or minimum impact begins to emerge.

Definition of Teaching Techniques

Since the purpose of this study was to determine the impact of various teaching techniques or methods on science achievement, it was important to define what the techniques were. This was necessary in order to select appropriate studies and clearly communicate results. It was also necessary because another group in the Colorado project was analyzing what was referred to as instructional systems. An initial definition provided by the project staff was as follows:

> Teaching methods are thought of as narrower, less encompassing than instructional systems. Whereas the latter might plausibly guide a great many decisions about the organization and conduct of teaching a science course, teaching methods refer to more limited aspects of a teaching plan (e.g., the method of testing, type of questioning, wait-time and the like). Studies in which teaching methods are evaluated are typically of short duration and limited to one or two narrow topics.

This definition of teaching technique was used to define twelve categories of teaching techniques. They are audio-visual, focusing, grading, inquiry, manipulative, modified, presentation approach, questioning, teacher direction, testing, wait-time and miscellaneous. Each of these categories will be briefly discussed with representative examples included.

Audio-visual

Although the bulk of the media-based instruction was considered to be under the domain of instructional systems, some was limited enough in scope or duration to be appropriately considered as a teaching technique. Examples of experimental A-V teaching techniques that were compared with control methods are:

1. Films on a specific topic
2. Videotaped presentations
3. Audio-taped directions
4. Supplemental pictures, photos, or diagrams.

Focusing

Teaching techniques included in this category include those where something occurs to alert students to the objectives or intent of instruction. Focusing techniques may be employed before, during or after instruction General examples of these include:

1. Students provided with objectives
2. Objectives reinforced at different points during instruction
3. Various organizers of instruction.

Grading

Experimental techniques included here involve changes in the grading system that the researcher has reason to suspect may result in improved student performance. Specific examples are:

1. Use of pass/fail graining
2. Students assigning their own grades.

Inquiry-discovery

In general the teaching techniques that involved more student-centered less step-by-step teacher directed learning experience are included in this category. Very often the techniques were identified by the authors of a research report as being inquiry or discovery. In nearly all cases these techniques are compared with a control method identified as being "traditional," "expository" or "conventional." Examples of the techniques used are:

1. Inquiry lessons
2. Guided discoveries
3. Inductive laboratories.

Manipulative

Students operate, handle or in some way work or practice with physical objects as part of the instructional process. Generally a single device or kind of manipulation is involved in this group of techniques. Examples are:

1. Operation of a specific piece of apparatus
2. Physical practice of some skill
3. Sketching or drawing
4. Constructing something.

Modified

Studies in which a researcher changes a single portion of instruction to test for improved student achievement. In almost all cases the modification or revision is of instructional materials. Examples include:

1. Materials rewritten or annotated
2. Directions presented other than by written word
3. Change in laboratory equipment.

Presentation Mode

This broad category of techniques refers to the means of instruction where several changes in material have taken place, the setting of instruction is different or student approach or introduction to a topic or teaching arrangements differ from what is considered a more traditional method. Representative examples are:

1. Field trips
2. Group discussions
3. Individual or self-paced lessons
4. Games--simulations
5. Team teaching.

Questioning

Teaching techniques that involve the use of varying levels or position of questions in instruction belong to this group. Examples of specific techniques found here are:

1. Questions inserted in a film
2. Knowledge and comprehension level questions at the start of a unit
3. Questions before, during, or after an assigned reading
4. Use of high level questions.

Teacher Direction

Variations in the extent to which the learning task was spelled out for the student typified teaching techniques classified under teacher direction. Specific illustrations include:

1. Students conduct experiments or activities given only sketchy direction
2. Students select objectives and assumes responsibility for learning
3. Indirect instruction.

Testing

Techniques where tests were used in various ways with a veiw toward improved student achievement. Usually this involved a change in the frequency of testing, the purpose of testing, or the level of the test items. The use of feedback is also included here. Examples are:

1. Formative testing
2. Immediate or explanatory feedback
3. Diagnostic testing and remediation
4. Optional retesting
5. Testing to mastery.

Wait-time

This category included studies that used increased duration of wait-time. These can be identified as:

1. Long vs. short wait-time
2. Added pauses at key response points.

Miscellaneous

They are all the other teaching techniques not classifiable into any of the previous categories. Examples include:

1. Students performed extra experiments related to the topic of instruction on their own time.
2. Students viewed a film more than one time.

Clearly these categories are by no means an absolute system for classifying teaching techniques. Further it is evident that while some of the categories established are clearly defined, others do not lend themselves to precise specification. These categories were formulated and studies were placed in them after all reports had been coded. This allowed for careful consideration of each category and study in terms of all the documents reviewed. They represent the variety of means researchers have used to bolster science achievement by altering some aspects of the instructional situation. Altered teacher behaviors, student actions and responsibilities, classroom materials and equipment, time, and testing are all included.

Procedures

Selection of Studies

The literature base searched for studies relating teaching techniques and science achievement included microfilmed dissertations, ERIC documents and reports, and the periodical literature. Studies selected for possible coding from these sources first had to have titles that implied they dealt with what would be considered teaching techniques.

The basis for this judgment was the definition of teaching techniques or methods provided by the meta-analysis project steering committee and further clarified through discussions at a training session held for all persons involved in the meta-analysis project. The contents of each study thus slected were then examined to confirm that the study was relevant.

The sample of studies that were ultimately coded are further described by the following items:

1. Age of Subjects -- The studies used mainly included students in grades 6 through college. It was originally intended that studies with students in grades kindergarten through college would be involved. Part way through the coding process the decision was made to limit the sample to only those investigations using subjects in the 6-college range. This was necessary to maintain the total coding task to manageable proportions.

2. Geography -- Studies were limited to those written in English and reported in the United States.

3. Year of Publication -- No study published earlier than 1949 was used.

4. Control Group Used -- A control or contrast group identifiable as being traditional or conventional was necessary.

5. Sufficient Data -- Enough data was included in the report so that an effect size could be calculated and identified as being positive or negative.

Notes were kept on the reasons individual studies were rejected. The predominant reason for disqualification was that upon examination of the contents of studies it became apparent that they did not deal with experimental teaching techniques as earlier defined. The second most common reason for rejection was that not enough information was provided to allow for effect size calculation. Rather complete means have been developed (see Glass, McGaw, White, & Smith, 1980) to determine effect sizes even when treatment and control means and standard deviations are not available. For example, given values of $\underline{t}$ and sample sizes it is possible to determine an effect size (see Appendix A for effect size calculation formulas). But even with these techniques insufficient information would still sometimes halt the meta-analysis process. Studies were also rejected because subjects were outside the 6-college age range or because there was no control or contrast group.

Data Sources

Research studies were examined that came from the following documents:

1. Science education doctoral dissertations--these included all dissertations available on microfilm from the Ohio State ERIC Center that related teaching techniques and science achievement.
2. ERIC documents and reports--these were identified by a computer search of the ERIC data base to identify reports of teaching techniques in science.
3. Journal articles--these included searching for relevant studies in all issues of the Journal of Research in Science Teaching (1963-1981), all issues of the Journal of College Science Teaching (1970-1981), all issues of Science Education from 1970 to 1981.

There are certainly additional sources of research reports relevant to this study. But the sources examined should provide the bulk of the reports available and allow reasonable inferences to be made about the aggregate effect of various teaching strategies on science achievement.

Coding the Studies

In order to insure that important variables from each study were examined and recorded in a consistent way, a suitable coding form had to be developed. This involved identifying potentially useful variables and incorporating them into a concise and easy to use format.

Development of a coding form began at the training session in Colorado where discussion generated a number of the variables that were incorporated. These variables were then categorized and arrayed into a convenient format to produce the initial coding instrument. Revisions to the form occurred as the actual coding process proceeded. These modifications were made to best accommodate the studies being coded.

A total of 76 variables were included on the final coding form and classified into major categories. Each of these categories is identified and briefly discussed. A copy of the coding form and information about each variable are included in this report (see Appendix B).

1. Report ID (3 variables). This group of variables provides a document number and identifies the coder.

2. Study Data (4 variables). Variables here are used to account for single or multiple treatments and measures. The year and form of the study are also identified.

3. Student Data (17 variables). Characteristics of study subjects such as grade level, SES, and number of students involved are documented by this group of variables.

4. Teacher Data (13 variables). Teacher characteristics such as gender, age, educational background and experience teaching are part of this category.

5. Context Characteristics (3 variables). The size of the classes and the schools involved as well as the types of communities are accounted for here.

6. Design Characteristics (6 variables). Experimental design considerations such as means of assignment of subjects and teachers are observed.

7. Treatment (16 variables). The specific types of teaching techniques and the roles of the teachers and students as well as the duration of the treatment are among the variables that are parts of this category.

8. Outcome Characteristics (4 variables). Variables considered as outcome characteristics involve the kinds of measures used and the reliability and reactivity of each.

9. Effect Size Calculation (10 variables). The kind of data used to calculate effect size and the study effect sizes are included in this group of variables.

The Coding Process

The first group of studies coded were the microfilmed dissertations. These were selected by Colorado on the basis of title and provided by the ERIC Center at Ohio State. The contents of each of the more than 300 dissertations were examined locally to determine if they actually dealt with teaching techniques as earlier defined. The studies were further screened to include only those using appropriately aged subjects, a traditional or conventional control group, and sufficient data to calculate effect size.

The second group of studies coded were ERIC documents. Abstracts of some 2,000 ERIC available science studies were provided by Colorado. These were reviewed to identify those that appeared relevant to the teaching techniques question. Copies of the studies thus selected were obtained locally on microfiche and screened like the dissertations to determine which ones would actually be coded.

Finally on an issue-by-issue basis the journals were scanned. These were all issues of the Journal of Research in Science Teaching, the Journal of College Science Teaching and Science Education from 1970 to the present. Studies were selected for coding in the same manner used with dissertations and ERIC documents.

Once it had been determined that a particular study was useable it next had to be coded. This involved reading the study and locating or determining values for as many of each of the 76 variables specified on the coding form as possible. In cases such as student grade level, or measure reliability, variable values were stated in the study. Values for other variables, like SES, were inferred when sufficient evidence was available. In the case of study effect size, the value had to be calculated in every instance.

In this meta-analysis study effect size is a standard measure of the difference between an experimental teaching technique and a traditional method. The vast majority of effect sizes were computed using means and standard deviations. To compute effect size when comparing an experimental teaching technique to a control method, the mean of the control group is subtracted from the mean of the experimental group and then divided by the standard deviation of the control group. An effect size is designated as negative when the mean score of the experimental group is lower than that of the control group. A single study results in multiple effect sizes when there is more than one experimental treatment or when there are two or more post measures.

Data from the coding forms were key punched (see Appendix C for the layout of the 76 study variables on the computer cards) and analyzed using the Statistical Package for the Social Sciences (Nie, et al., 1975) on an IBM 360 computer. The analysis included descriptive statistics for all continuous and categorical variables, categorizing some variables originally coded as continuous, and crosstabulating a number of study characteristics (e.g., grade level of subjects) with effect size.

Results

The computer analysis of the 76 variables considered in the study produced the results that follow. Note throughout this portion of the report that many of the variables are not mentioned primarily because there was limited information available in the research documents. For example, the educational background of teachers was reported in fewer than 30% of the studies and is not discussed here.

Descriptive Data

Research studies of the effects of various teaching strategies were examined that covered approximately the last 30 years. Figure 1 shows the distribution of effect sizes associated with different spans of years. The early 1970's produced the largest group of effect sizes (40% of the total).

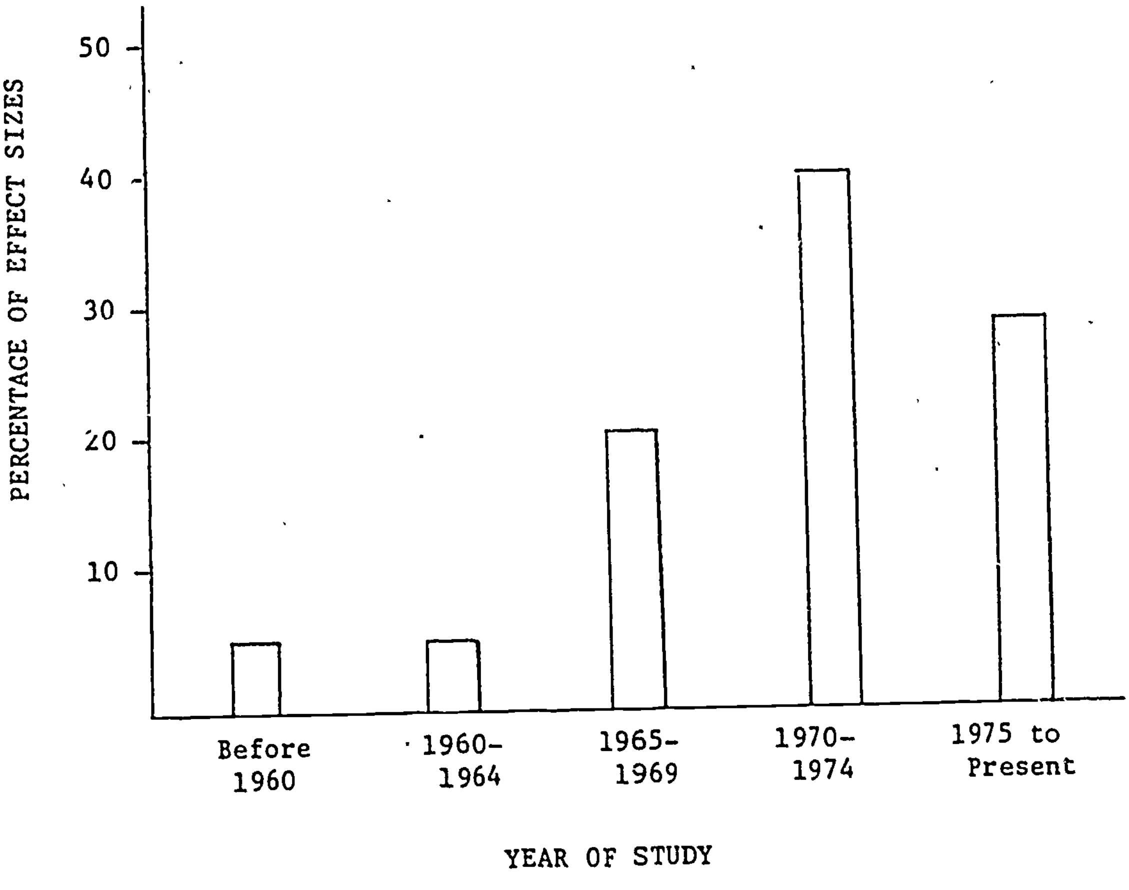


Figure 1. Percentage of effect sizes represented by different time periods.

The studies selected for the meta-analysis were conducted with students primarily from grade 5 through the early college years. Figure 2 shows the percentage of effect sizes associated with each span of grade levels.

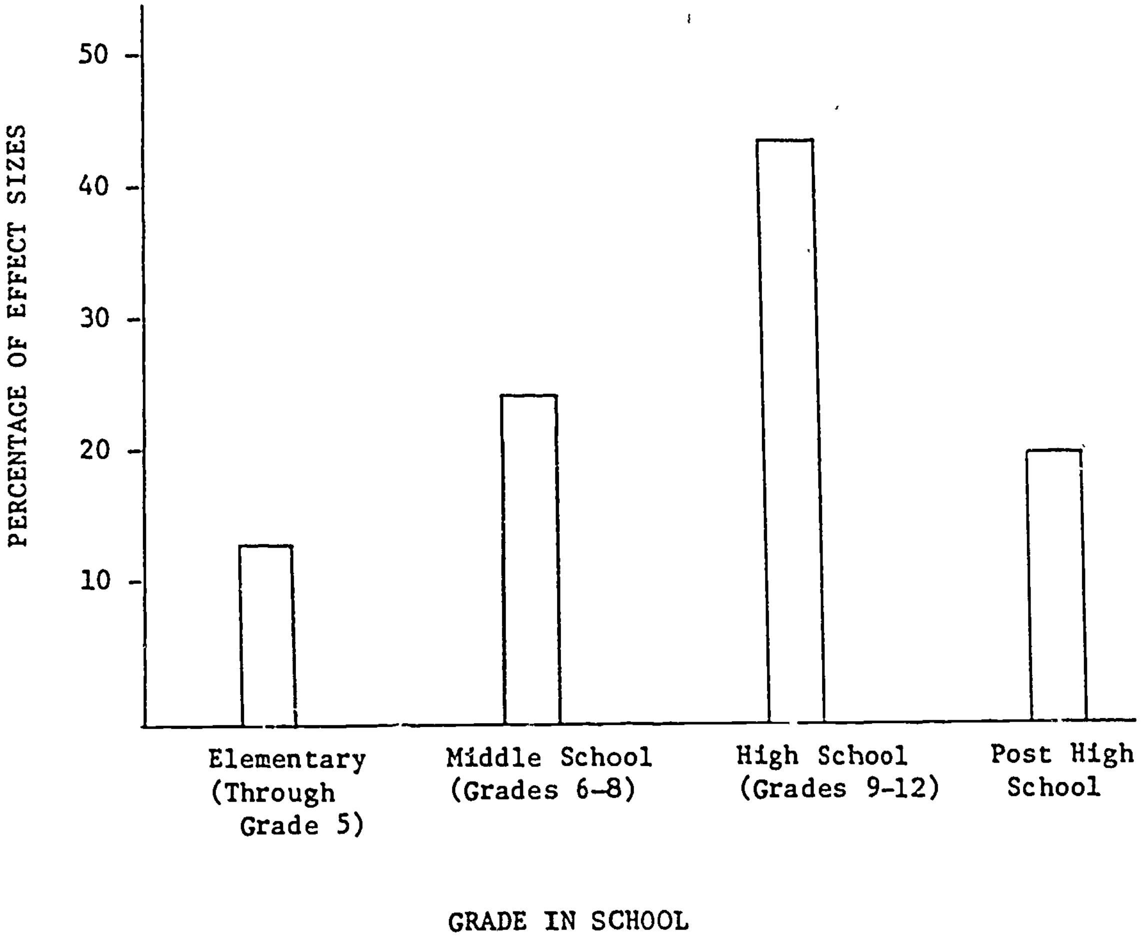


Figure 2. Percentage of effect sizes represented by students at different grade levels.

The traditional areas of study in science are represented among the research reports included in the teaching strategies meta-analysis. Figure 3 shows the percentage of effect sizes for each science subject area.

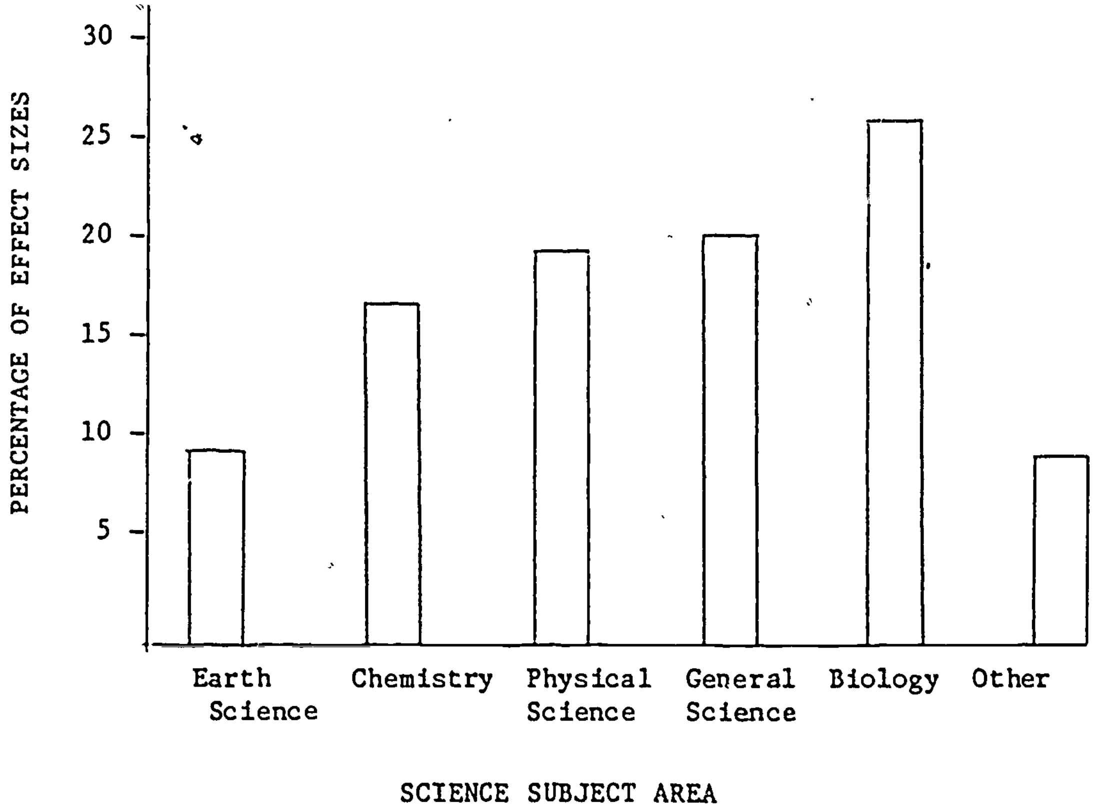


Figure 3. Percentage of effect sizes represented by different subject areas.

Outcome measures used in the studies of teaching strategies ranged from traditional cognitive tests to interview techniques. Figure 4 shows the distribution of outcome measures.

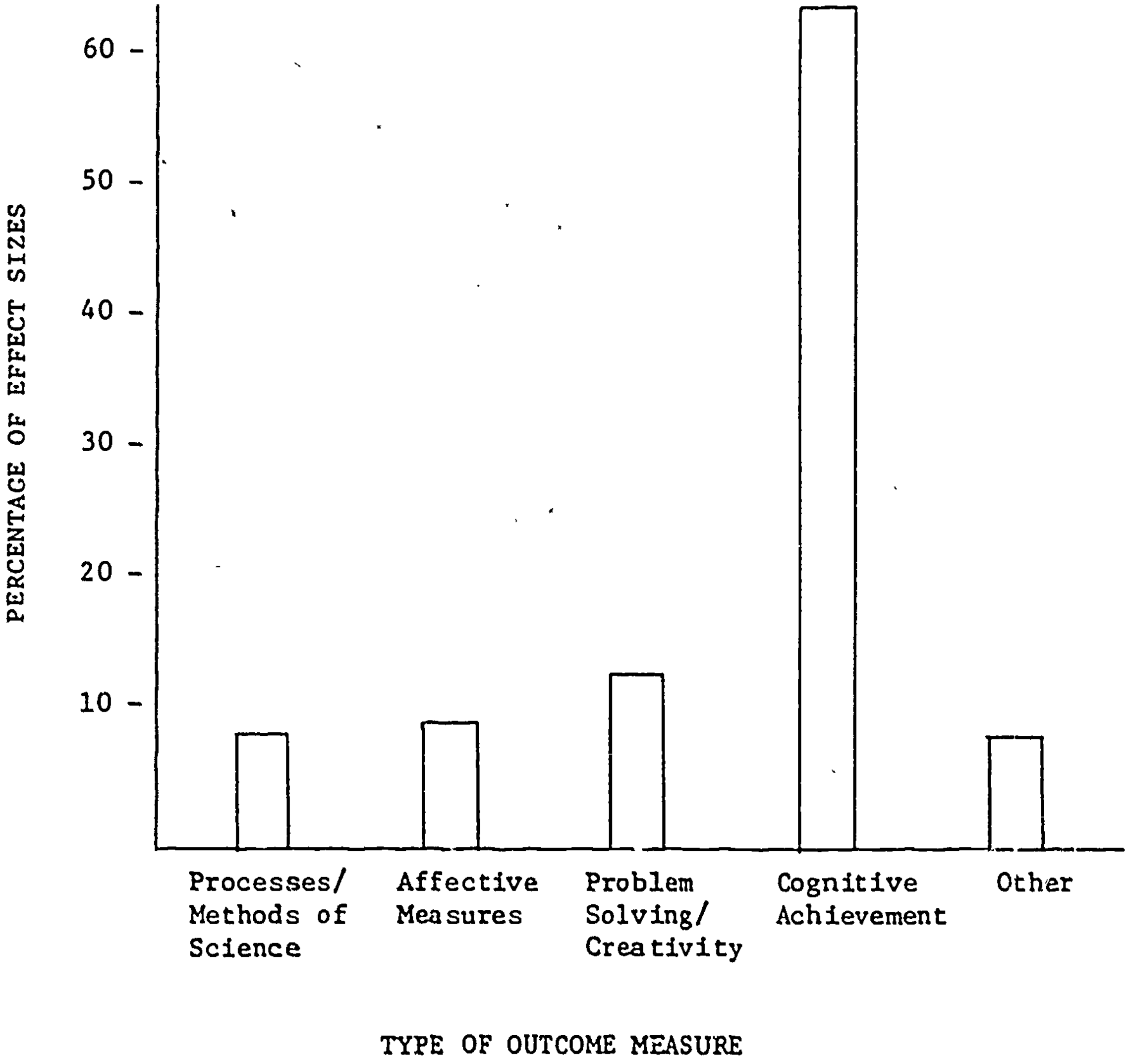


Figure 4. Number of effect sizes associated with different types of outcome measures.

Additional data that provide information about the contexts in which the selected studies were conducted are given below.

1. The effect sizes came from three types of reports:

Dissertations	56%
Journal Articles	26%
Unpublished Papers	18%

2. The research studies were conducted in different types of community settings.

Rural-town	17%
Suburban	35%
Urban	20%
Not Classified	29%

3. The instruments and methods used to measure outcomes were distributed as follows:

Instruments developed especially for the study	38%
Published instruments	31%
Regular classroom tests	20%
Observations and interviews	8%
Other	3%

4. The number of subjects included in the various studies were distributed as follows:

50 or fewer subjects	12%
51 - 99	28%
100 - 199	30%
200 or more subjects	27%
Unknown	3%

5. The number of teachers involved in the studies are as follows:

1 or 2 teachers	28%
3 - 8	21%
9 or more	9%
Unknown	42%

6. The teaching strategies were used in classes of various size distributed as follows:

Fewer than 15 students	8%
15 - 24	29%
25 - 34	28%
35 or more students	9%
Unknown	26%

7. The studies in which the teaching strategies were used were conducted over varying lengths of time.

2 hours or less	15%
3 - 10	19%
11 - 20	4%
More than 20 hours	32%
Unknown	30%

8. The reliability of the criterion measures used were distributed as follows:

.69 or less	6%
.70 - .89	36%
.90 - 1.00	10%
No information given	48%

Effect Size Data

A total of 160 studies were coded resulting in 411 effect sizes. The overall mean effect size on all outcome variables in this analysis is .34 (for all teaching strategies). When divided into two categories, cognitive and other, the results were very similar ($\bar{\Delta}=.35$ and $\bar{\Delta}=.30$ respectively) The average impact of using one of the teaching strategies analyzed in this report, therefore, was to increase achievement by about one-third of a standard deviation. In terms of percentiles, the mean effect of using the teaching strategies was to increase scores by about 13%.

Not all teaching strategies had the same impact on achievement. Table 1 gives the mean effect size for the 12 categories of teaching strategies used in this analysis. More confidence can be placed in some of these compared to others because of the number of effect sizes represented by each mean score.

Mean effect sizes were calculated for studies conducted with students of different grade levels. Table 2 provides mean effect size information for the four categories of grade levels used.

This meta-analysis of the effects of teaching strategies provided some additional data concerning learning in classes of different size. Table 3 presents findings similar to those of Glass and Smith (1979) in their study of class size and achievement. The largest effects of the different teaching strategies are associated with the smallest class size.

Different mean effect sizes were found in the different academic areas represented by the studies in the analysis (see Table 4).

Mean Effect Sizes Obtained for Cognitive and Other Outcomes Obtained Using Different Teaching Strategies

Type of Strategy	Cognitive* $\bar{X}$	Cognitive* SD	Cognitive* n	Other** $\bar{X}$	Other** SD	Other** n	Total $\bar{X}$	Total SD	Total n	% of All Cases
Wait-Time	.53	.02	2	1.27	.00	2	.90	.43	4	1
Focusing	.48	.90	25	1.37	.63	3	.57	.91	28	7
Manipulative	.56	.64	24	--	--	--	.56	.64	24	6
Modified	.55	.45	20	.27	.34	2	.52	.45	22	5
Questioning	.56	.37	11	.07	.06	2	.48	.39	13	3
Inquiry-Discovery	.41	.87	38	.15	.29	20	.32	.73	58	15
Testing	.37	.49	33	.14	.34	11	.32	.46	44	11
Presentation Mode	.24	.54	77	.29	.62	26	.26	.56	103	26
Teacher Direction	.18	.54	28	.32	.83	17	.23	.66	45	11
Audio-Visual Methods	.16	.49	30	.33	.35	3	.18	.48	33	8
Grading	-.13	139	13	-.40	.00	1	-.15	.38	14	4
Miscellaneous	.53	.24	8	.23	.19	4	.43	.26	12	3
Total	.35	.64	309	.30	.61	91	.34	.63	400	100

---

*The Cognitive category includes low and high level outcomes, general achievement, and problem solving.

**The "Other" category includes critical thinking, creativity, logical thinking and affective measures.

Table 2

Mean Effect Sizes Obtained with Students at Different Grade Levels

Grade Level of Student	Mean Effect Size	SD	Number of Cases	% of Cases
Elementary (through Grade 5)	.36	.71	50	13
Middle School (Grades 6-8)	.30	.74	93	24
High School (Grades 9-12)	.29	.53	164	43
Post High School	.42	.70	77	20

Table 3

Mean Effect Sizes Obtained in Classes of Different Size

Size of Class	Mean Effect Size	SD	Number of Cases	% of Cases
Fewer than 15 Students	.74	.86	32	11
15 - 24	.37	.60	119	39
25 - 34	.23	.46	114	37
35 or more Students	.23	.57	38	13

Table 4

Mean Effect Sizes Obtained in Studies Focusing on Different Academic Areas in Science

Area of Study	Mean Effect Size	SD	Number of Cases	% of Cases
Physical Science	.55	.81	78	19
General Science	.35	.65	81	20
Biology	.25	.55	105	26
Chemistry	.22	.53	68	17
Earth Science	.12	.53	36	9
Other	.52	.54	35	9

Another means of examining mean effect sizes is by the source of the literature report. Table 5 gives mean effect sizes for the three types of reports examined in this analysis.

Table 5

Mean Effect Sizes for Various Types of Literature Reports

Type of Report	Mean Effect Size	*SD*	Number of Cases	% of Cases
Journal Articles	.41	.67	105	26
Dissertations	.32	.66	230	56
Unpublished (ERIC Documents)	.30	.51	74	18

The studies reviewed for this analysis included widely differing numbers of students. Table 6 provides mean effect sizes associated with the studies of various size.

Table 6

Mean Effect Sizes Obtained in Studies with Different Number of Students Involved

Number of Subjects in Study	Mean Effect Size	SD	Number of Cases	% of Cases
0 - 50	.66	.90	49	12
51 - 99	.41	.71	115	29
100 - 199	.35	.53	125	31
200 or more	.09	.38	110	28

The average effect sizes associated with studies using different numbers of teachers is shown in Table 7.

Table 7

Mean Effect Sizes for Studies Involving
Different Numbers of Teachers

Number of Teachers	Mean Effect Size	SD	Number of Cases	% of Cases
1 - 2	.41	.70	116	28
3 - 8	.35	.56	86	21
9 or more	.20	.30	36	9
Unknown			173	42

The studies analyzed were conducted over widely differing amounts of time. Some were done in only a class period or two while others lasted for several months. The information in Table 8 shows the mean effect sizes associated with four categories of study time.

Table 8

Mean Effect Sizes for Studies Conducted for Different Amounts of Time

Duration of Study (Hours)	Mean Effect Size	SD	Number of Cases	% of Cases
0 - 2	.44	.84	63	15
3 - 10	.43	.71	77	19
11 - 20	.20	.36	16	4
More than 20	.33	.57	132	32
Unknown			123	30

Interpretation and Implications

What conclusions should be drawn from an integration of the research on teaching strategies in science that produces an overall effect size of about one-third of a standard deviation (.336)? Are alterations in such things as teacher questions or directions, student activities, classroom materials, tests or grading practices worth the effort when they result in student scores that are on the average 13 percentile points higher than in the unaltered classes? These questions are probably unanswerable and perhaps even unimportant unless one is concerned about the overall impact of innovations. Most often teachers, teacher educators, researchers, or instructional developers have interest in a certain instructional strategy. Thus their concern is with the impact of a specific teaching strategy and not the effect of all teaching strategies. Even the clumping of strategies into 12 categories as has been done in this report makes it difficult to determine the impact of a teaching strategy such as "providing students with instructional objectives" because it is lumped with all other Focusing strategies.

The picture provided by this meta-analysis of teaching strategies is a macroscopic view. It provides evidence on the general, overall impact of a category of strategies but does not give fine, detailed microscopic information on a particular strategy.

The information in Table 1 shows a range of effect sizes from .90 (Wait-Time) to -.15 (Grading). Recent reviews of instructional research (e.g., Rosenshine, 1979) have concluded that direct teaching strategies have greater impact than indirect ones. Is that conclusion supported by this analysis? There appears to be some support among the strategies with relatively large effect sizes; higher than the mean are Focusing and Questioning and among those with relatively small effect sizes are Inquiry-Discovery and Teacher Direction. Wait-Time strategies have the largest impact but they also account for the fewest number of studies reviewed in any category.

The effect sizes associated with classes of different size (see Table 3) should provide strong evidence for policy makers who advocate smaller classes. In this case the accumulated research results confirm teachers contentions.

Research methodologists can glean items of some interest from this report. Studies involving the fewest number of subjects produced the largest effect sizes (Table 6). Essentially the same information is obtained by examining the number of teachers in a study and the related effect sizes (Table 7). Again, the smaller number of teachers involved the larger the effect size. Both the number of subjects and number of teachers may have much to do with faithful implementation of a strategy. Treatment fidelity may suffer when large numbers of students and correspondingly large numbers of teachers are involved.

Another point of methodological interest is seen in the data on duration of a study (Table 8). Short studies produced larger impacts (larger mean effect sizes) than long studies. This may also relate to treatment fidelity wherein the control over a strategy may lessen as a study stretches on.

There seems to be little to say about the effect sizes associated with different academic areas (Table 4) or grade level of subjects (Table 2). A pattern in the effect sizes that begins in the grade level data (elementary → middle school → high school) is broken with the college students. For the academic areas the more general topics seem to yield higher effect sizes than specialized subject areas such as biology, chemistry, or earth science.

Critics or even advocates of instructional strategy research may feel that the overall impact of the various strategies is somewhat small. Several points should be made, however, to show that the aggregate score may obscure much information about impact. Among the studies examined, the range of the effect sizes was nearly 6 standard deviation units. The largest effect size was 3.58 and the smallest was -2.10. A useful analysis of a particular teaching strategy might be to identify features of studies (either design, treatment, or context) associated with large and small impact. It may be possible to refine a teaching strategy by emphasizing features associated with large impact and minimizing or dropping features associated with little effect. Subsequent research could then determine if the adjustments were advantageous. It is interesting to imagine how several strategies, none of which has an overwhelming impact, might influence achievement if used in concert. Consider classes in which Focusing strategies (effect size = .57), Questioning (effect size = .48), and Testing (effect size = .32) were combined by teachers. The overall influence might not be the simple sum of the individual contributions but a combined influence would be expected. This engineering of teaching strategies to optimize achivement would seem to have much promise.

REFERENCES

Glass, G. Primary, secondary and meta-analysis of research. Educational Researcher, 1976, 5, 3-8.

Glass, G. Integrating findings: The meta analysis of research. Review of Research in Education, 1978, 5, 351-379.

Glass, G., McGaw, B., & Smith, M. Meta Analysis in Social Research. Beverly Hills, California. Sage Publications, 1981.

Glass, G., & Smith, M. Meta analysis of research on the relationship of class-size and achievement. Evaluations and Policy Analysis, 1979, 1, 2-16.

Nie., N., Hull, C., Jenkins, J., Steinbrenner, K., & Bent, D. Statistical package for the social sciences. New York: McGraw-Hill, 1975.

Rosenshine, B. Content, time, and direct instruction. In P. Peterson, & H. Walberg (Eds.), Research on teaching. Berkeley, CA: McCutchan, 1979.

Willson, V. A meta-analysis of the relationship between science achievement and science attitude: Kindergarten through college. A paper presented at the National Association for Research in Science Teaching annual meeting, Grossinger, New York, April, 1981.

Yeany, R., & Miller, P. The effects of diagnostic/remedial instruction on science learning: A meta-analysis. Journal of Research in Science Teaching, 1982, 19 (in press).

The effect of inquiry teaching and advance organizers upon student outcomes in science education: A meta-analysis of selected research studies

Gerald W. Lott
Institute for Research on Teaching
Michigan State University

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The purpose of the research was to determine the relationship between variations in the nature and structure of instructional content and outcome variables across the relevant experimental studies. Included here are the comparison of the inductive vs deductive approach (Shulman & Keisler, 1966), and the use of advance organizers (Ausubel, 1960, 1963). Of several areas dealing with the nature and structure of instructional content, only these areas of research provided a data base of sufficient depth to justify a meta-analysis. Among the topics considered but not included here are behavioral objectives, kinetic structure, mathemagenic behavior, curriculum scope and curriculum organization. Thus the data base for this study accounts for the coding of 128 characteristics for 39 studies selected from the 72 coded as part of the larger study. The larger study in turn is but one part of ~~a~~ the broader project to integrate science education research.

The author would like to acknowledge the valuable suggestions provided by Edward L. Smith and Lee S. Shulman, Institute for Research on Teaching, Michigan State University, and Carol Blumberg, Department of Educational Studies, University of Delaware during the final revision of the coding sheet.

Various normative arguments have led to the development and execution of numerous empirical studies. The guiding assumption for the current analysis has been that these studies could provide information for further research through quantitative analysis of their characteristics. Relative patterns which exist among appropriate variables of each study could be revealed through the utilization of Meta-Analysis.

This process involves viewing the studies as the units on which measurements are taken with the variables being the coded study characteristics and effect sizes. The coding variables included 57 which were concerned with features of the treatment while 12 were concerned with outcome attributes. Aspects such as methodology, sample characteristics, and instructional experiences were examined quantitatively in terms of their relationships to the treatment effects through the use of a common metric for all studies as defined by Glass (1978).

METHODOLOGY

Selection of Studies

Studies included in this analysis were selected from the one-hundred and fifty-one studies provided for possible inclusion by the Science Meta-Analysis Project. These studies were then examined to (1) determine their relevance to the broader research question and (2) ascertain the availability of the necessary data for effect size calculation. Those studies having means and standard deviations were coded first, while other studies requiring more extensive calculations or which had minimal data reported were set aside for coding if time permitted. Several journal reported studies were coded based upon a limited search of articles in the Journal for Research in Science Teaching for 1977 - 1980. However, major emphasis was given to the coding of dissertation studies. This examination resulted in a collection of 105 studies found to be relevant to the topic and having sufficient data for the calculation of effect sizes.

Studies analyzed and reported in this report spanned the period from 1957 through 1980. Most of the studies were conducted during the 1969 through 1973 period. The majority of the studies used in the meta-analysis reported here were dissertations with 33 studies (85%) being nonpublished doctoral dissertations and 6 studies (15%) being articles published in professional journals. All dissertations received were completed prior to 1977. It is estimated, based upon a review of the references given in those studies coded as well as a survey of reviews of science education research, that this analysis represents approximately

35% of the advance organizer research, and 25% of the inductive-deductive research.

The systematic analysis of these studies resulted in the coding and calculation for 424 effect sizes. A separate effect size was calculated and study characteristics coded for each distinct outcome variable within each study. An average effect size was calculated for each instance where a particular outcome characteristic value would be used several times within a comparison for a study.

Identifying and Coding Variables

To make full use of statistical methods in describing and communicating study findings and accounting for their variance, the characteristics of the subjects, teachers, context, design, treatment, and assessment for each study were expressed in quantitative terms. Some of the features and the nature of the assessment procedures used an ordinal scale while others involved a nominal scale based upon indicator variables.

The development of the coding form involved the preliminary coding of several articles. In addition, a survey of several reviews of science education research as well as learning and cognition reviews was made to ascertain the important characteristics to be coded. This resulted in the adoption of several classifications based upon categories proposed in the literature. The following provides a brief overview of the characteristics which were coded. A description of the conventions used for many of the coding sheet items as well as the coding sheet is provided in ~~Anderson, et al. (Note 1).~~ the appendices of this part.

The study identification variables were used to distinguish studies as well as multiple effect size codings within single studies. The study code identified each individual study. Each comparison within a study was given a code while within each comparison a separate code was given. Thus, those studies which compared more than one treatment, varied the sample characteristics, or used more than one outcome measure for any comparison were given distinct codes.

The student characteristics variables were intended to delineate various important features of the samples used. Teacher characteristics variables provided background information concerning the individuals presenting the instructional treatment. Characteristics of the context were intended to provide necessary information concerning the environment in which the study was conducted.

Coding for design characteristics included the methods for assigning students and teachers to the treatments as well as the unit of analysis and experimental design used. The coding for the internal validity of each study followed the convention that intact and highly dissimilar samples, based upon ability or socio-economic level, were classified as having low internal validity. Those found to have intact or randomly selected classrooms with similar characteristics were coded as medium, while those studies which involved complete random selection of subjects and had low mortality were coded as high. The coding for the type of study followed the system proposed by Campbell and Stanley (1963).

Treatment characteristics included such aspects as preinstructional strategies, the inquiry orientation of instructional tasks, the characteristics of the learning tasks as well as the content, the type of instructional techniques. The preinstructional stragegies included the coding for the type of advance organizer used. The distinctions used for the coding of the level of inquiry were based upon those suggested by Shulman and Tamir (1973). The coding for the characteristics of the learning tasks involved items concerned with the kinds of activities (Johnson, Rhodes, and Rumery, 1974). Categories for the structure of content were based upon those suggested by Haggis and Adey (1979) while those used for the characteristics of the questions asked as part of instructional tasks were based upon those proposed by Bloom (1956) for the level cognitive reasoning and by Johnson, Rhodes, and Rumery (1974) for the level of generality.

The coding for the outcome characteristics included categories such as the intent of the assessment, the domain orientation, the type and method of measurement, the reactivity, and the reliability. The convention for coding the intent of the assessment was based upon the novelty of the context (Johnson, Rhodes, and Rumery, 1974); i.e. whether it was to assess the acquisition of knowledge involving identical or similar aspects, or whether it was to assess transfer to related or new situations. The domain of orientation distinguished between cognitive, affective, and behavioral forms of assessment. The convention for reactivity involved the specification of whether judgments were objective or subjective.

In general, most studies were very limited in the description of study characteristics and thus several of the variables were never or seldom used. As a result nearly 25% of the variables were eliminated early in the data analysis process including almost half of the student characteristic variables and nearly all of the teacher characteristic variables. In addition, thirty percent of the treatment characteristic variables were eliminated due to limited usage.

Calculating Effect Sizes

The calculation of effect sizes for this study utiltized where possible the definition proposed by Glass (1978). Generally, if post-test means and standard deviations were provivded, this procedure was used. In other cases the appropriate approaches as presented by Glass, McGaw and Smith (1981 were utilized.

Analysis of Effect Size Data

The approach utilized in this study for effect size data analysis was within the exploratory data analysis paradigm. The delineation of the approach to the analysis of effect size data will consist of an elucjdation of the exploratory nature of the analysis, and the statistical approaches and sequence use.

It is argued that the data analysis procedures following data acquisition should be exploratory. The use of these procedures prior to the further application of inferential statistical methods can provide the data base for the formulation of "conjectures" and the resultant design of "experimental arrangements." The concept of conjecture is used in the sense proposed by Popper (1962) while experimental arrangement is a concept elaborated by Hanna (1966) which refers to the nature of the treatment variables. I am, however, using it in a broader sense to include the presage and process variables.

The intent is to formulate questions for further research and provide direction for research programs. Tukey (1980) has indicated that research questions should be formulated only after extensive exploration of the data base. Exploratory data analysis potentially can provide insights regarding features of the experimental arrangements. These discernments can clarify interrelationships and give direction to the science education research effort.

Leinhardt and Leinhardt (1980) suggest that exploratory data analysis is "an approach that illuminates rather than obscures the analysis of data and makes apparent rather than disguises analytic results" (p. 85). They later point out that:

> The philosophy is one in which the analyst's first task is viewed as discovery of evidence, not evaluation, and consequently the tools are designed to reveal unforeseen features rather than create a decision-analytic framework for judging the importance of expected features" (p. 149).

These quantitative techniques give direction to future research through descriptive analysis rather than providing a basis for confirmatory inferences. Thus, exploratory data analysis can provide "descriptive power" (Hanna, 1969), the essential characteristic of which is that the descriptive power of models which reflect additional information transmitted by the data. It is within this framework that exploratory data analysis can assist in developing further experimental arrangements with explanatory power.

The intent of the procedures described below is to expose the relationships between study characteristics and effect sizes using descriptive data. The statistical analysis began with the use of SPSS FREQUENCIES. The next step involved the use of SPSS CROSSTABS to organize the study characteristics data systematically in contingency tables for further analysis.

Since the principal task of this study is exploratory rather than confirmatory, the findings will be summarized using exploratory data analysis methodology (Tukey, 1977). The data displays were box-and-wisker plots, a technique used to display batches of data. The median as well as the upper and lower quartiles are calculated and the quartiles used to define, on a vertical axis in this study, the boundaries of a narrow rectangle. The length of the box is used to define "inner fences," while 1.5 times this distance defines "outer fences." These conventions are a modification of those proposed by Tukey (1977) and are based upon those used by McNeil (1977).

Analysis and Interpretation of Data

The guiding interest was determining the relationship between effect sizes and study characteristics. The intent was to analyze the population of effect sizes across characteristics within each of two research areas: Inductive vs Deductive, and Advance Organizers. The principal goal of this analysis was to determine the relationship between effect sizes and study characteristics within these selected research areas through a comparison between effect sizes across the levels of each descriptive variable for each of the two defined research variables. This approach includes the review of correlation coefficients, the examination of study design characteristics in relation to the effect size, and treatment characteristics in relation to effect size. All of these calculations were done using microcomputer programs based on those written by McNeil (1977).

Due to the limited number of studies for each of the two defined research variables a "dependent measure" approach was used in the analysis where each of the 424 effect sizes were treated statistically as an independent data point. Thus, any dependence of the data due to those studies yielding more than one effect size due to multiple but distinct outcome instruments or factorial designs providing multiple comparisons is not accounted for. However, a sample of study characteristics was selected and the median and hinges were calculated and found to differ by only .07 effect size for the "dependence approach."

The following sections explore the relationships between each study characteristic variable and the effect sizes for each research variable. This analysis will be limited to those cases where the crosstabulated cells for the research variable by study characteristic in question have a sufficient number of cases. To justify any discussion intended to lead to useful recommendations and conclusions, it was deemed necessary to have 10 or more effect sizes and 4 or more studies.

RESULTS OF DATA ANALYSIS

Inductive vs Deductive

This research variable, defined by the crosstabulation of those items specifying an inductive or deductive approach for the experimental and control group, can be characterized by the learning activities sequence. Educational experiences in which examples or observations were provided to students prior to formalizing generalizations were classified as inductive. Those studies where generalizations were formulated prior to any illustrative examples were characterized as deductive. The analysis of the data base found 212 effect sizes from 24 studies where the treatment or control group was coded as inductive or deductive.

Table 1 shows the mean effect size for inductive vs deductive teaching on several outcome measures. Effect sizes in favor of the inductive approach are labelled positive and those in favor of the deductive approach are designated negative. The composite of these several outcome measures has an effect size of .06. In the aggregate, there is essentially no difference between the two approaches.

A summary analysis of this data base across all study characteristics resulted in finding a median effect size of .02 and hinges of .33 and -.22. Effect sizes in favor of the inductive approach are designated positive and the deductive negative. Thus, the average student experiencing an inductive instructional approach did better than only 51% of the control group. It must be remembered, however, that approximately 60% of the studies used a level of inquiry only slightly different from the deductive approach. In addition, the spread between hinges indicates that 75% of those have an inductive approach did better than 42% of the control group.

Table 1

Mean Effect Sizes for Inductive vs. Deductive Teaching

on Several Outcome Measures

	$\tilde{\Delta}$	s	n
Knowledge	.02	.67	18
Application	-.10	.36	4
Process	.29	1.57	8
Problem Solving	-.01	.57	8
Composite	.06	.87	38

The following narrative will describe any differences in effect size based on particular study characteristics. This description will provide insights as to circumstances where an inductive approach could be expected to provide a more effective educational approach.

Several variables were not analyzed due to the lack of variation in study characteristics across studies within this research topic. These were features with insufficient data for comparative analysis but which provided added detail concerning the nature of treatments within this research category. All studies which described the grouping patterns did not have grouped subjects. The scope of content for the majority of effect sizes coded was disciplinary. Within this framework the organization of content was generally concept-oriented with some treatment comparisons using an organizational scheme involving topics. While there was somewhat more variation in the features concerning manipulative level, most of the treatments were characterized as having individual manipulation with objects. However, nearly half of the treatment comparisons utilized picture study. The majority of studies did not fully describe the mode of communicating knowledge. Those who did generally used the laboratory although nearly one-third used demonstration. Fifty-five percent of the outcome comparisons were concerned with the acquisition of knowledge by subjects. Of these studies 66% were characterized as assessing student performance on knowledge similar to that used in the instruction.

The correlation of data within this research topic resulted in some statistically significant correlations (where $p < .01$) Those with $r \geq .49$ and having important methodological and educational implications are shown in Table 2.

The following findings are evident in Table 2.

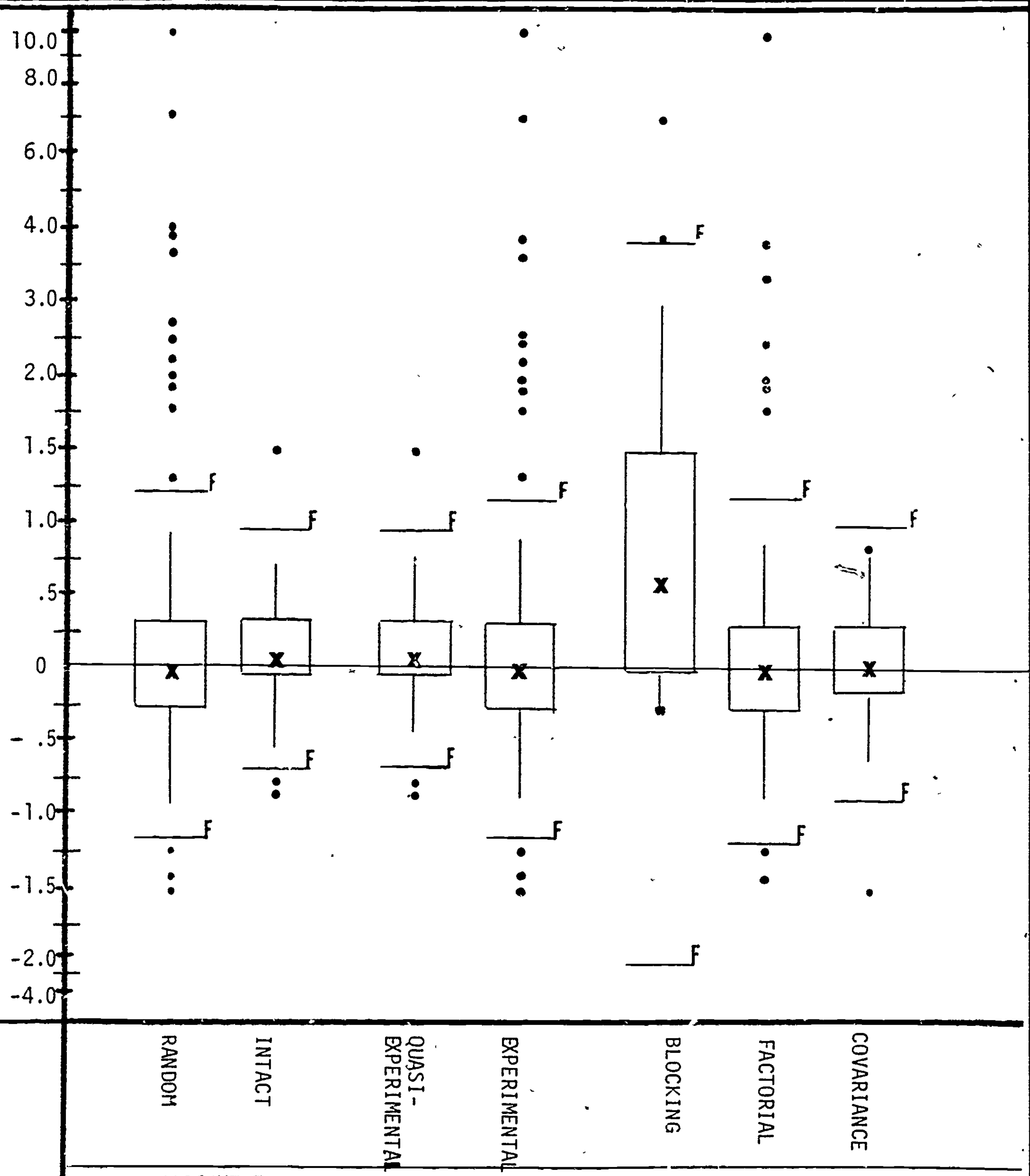
1. Those studies with samples having higher IQ ability had a more heterogeneous make-up.
2. Those studies conducted in a suburban environment had subjects from higher socio-economic status than those studies conducted in an urban area.
3. Those studies whose assignment of subjects was less experimental generally involved a longer duration. Experimental studies were usually of shorter duration. On the other hand, those studies using random selection of subjects generally involved more sessions.
4. Those studies classified as inductive were generally conducted in an urban environment and involved quasi-experimental designs with low internal validity.
5. Studies which utilized designs with highe internal validity in experimental framework generally involved the use of a more structured learning environment.
6. The studies with guided exploration utilized fewer sessions than those which were structured.
7. As might be expected, the deductive approaches were more structured than the inductive-oriented learning experiences.
8. The inductive approaches utilized a higher level of inquiry as opposed to those which were deductive.
9. Studies with higher-level inquiry involved less restrictive environments and greater access to manipulative activities than those studies having lower levels of inquiry.

10. Approaches in which subjects made judgments or organized elements into new patterns were inductive-oriented with a higher level of inquiry than those which required subjects to simply retrieve information.
11. Those studies aimed at developing or assessing concepts generally utilized the biological sciences for their content orientation.

The data analysis resulted in the observation of several differences which are important for future methodological decisions. These results are shown as graphical box-and-wisker plots in Figure 1, where fences are used to identify the outside and far out values of the effect sizes for each comparison. These comparisons were made to discern if the data required any separate analysis for those studies with a stronger design as a result of being noticeably different from the complete data base.

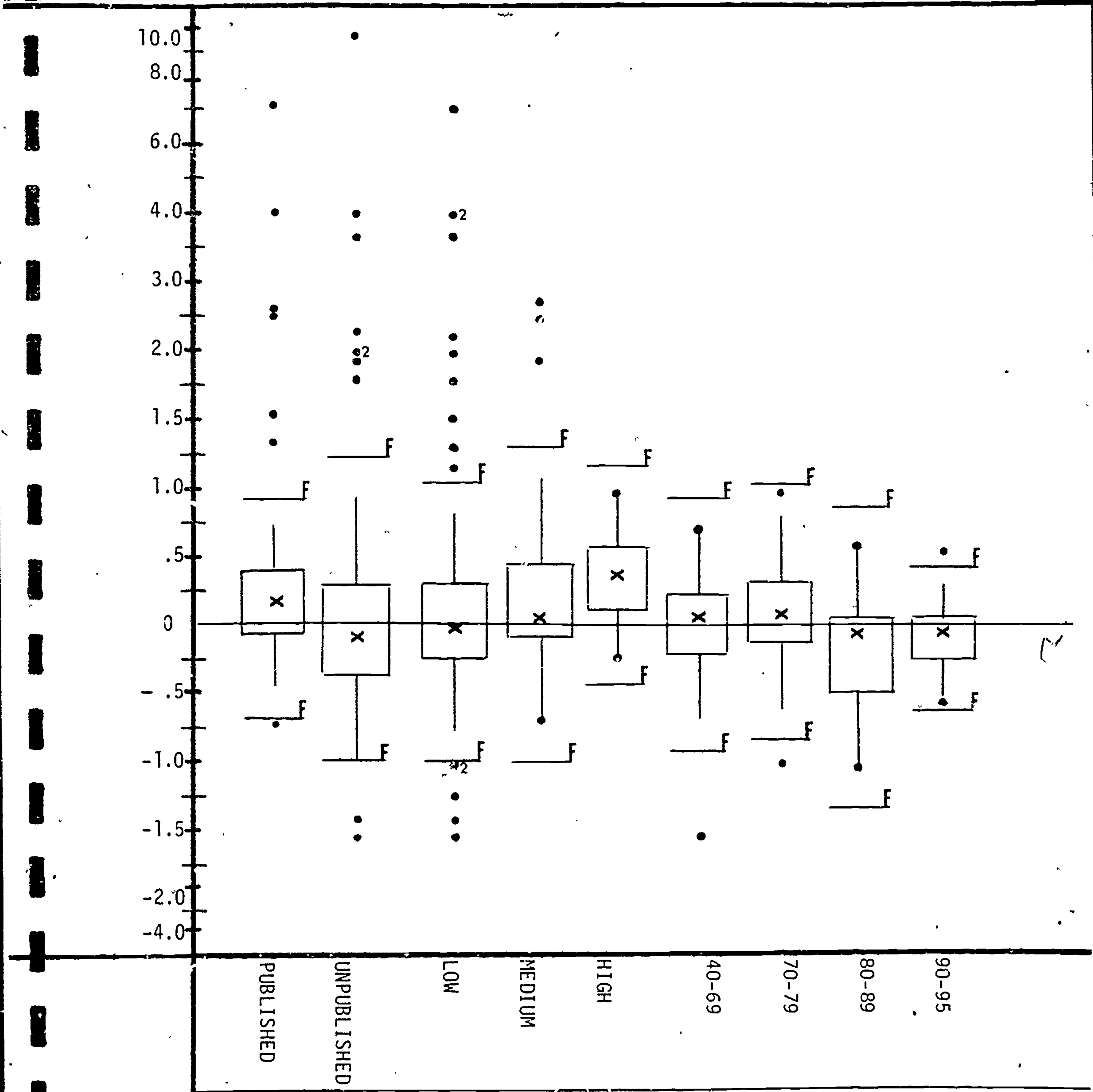
Figure 1

FIGURE 1 BOX-AND-WHISKER PLOTS FOR DESIGN CHARACTERISTICS, OUTCOME CHARACTERISTICS, AND SOURCE OF EFFECT SIZE



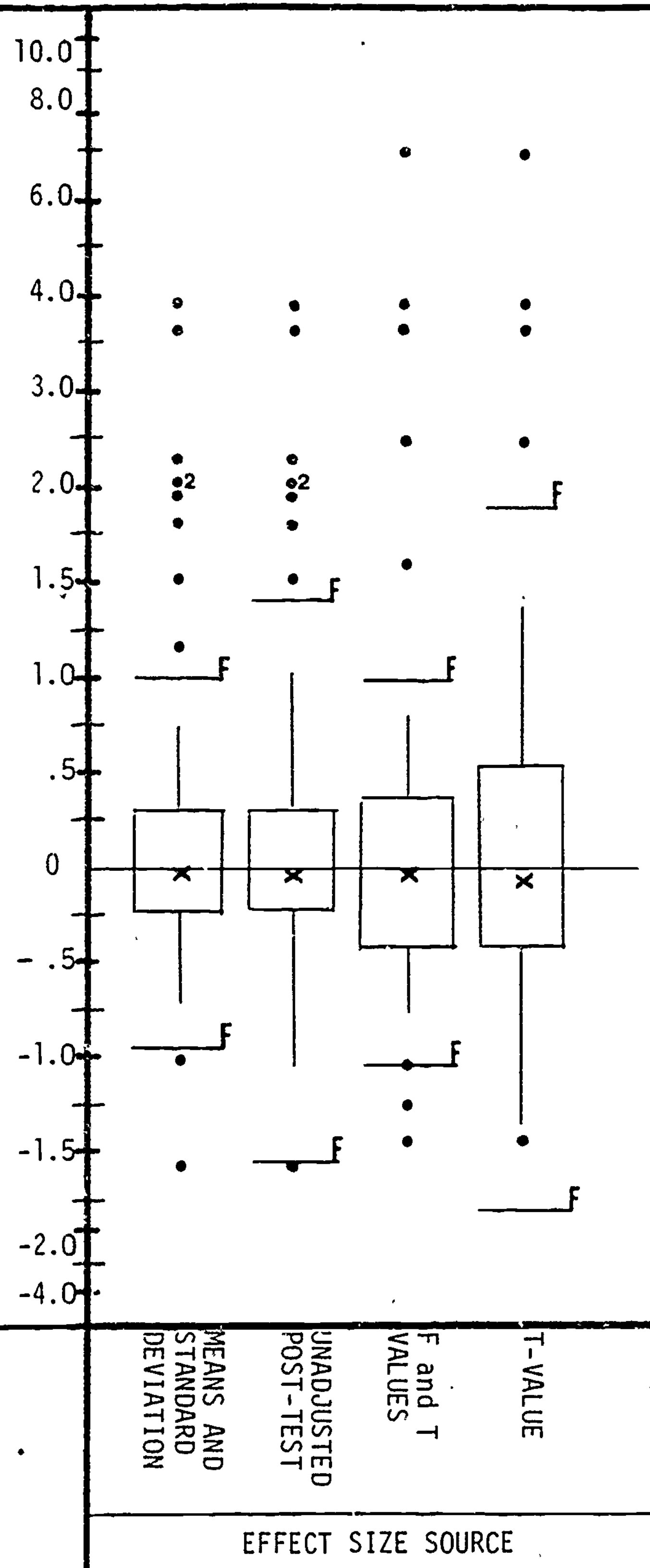
	ASSIGNMENT OF SUBJECTS		TYPE OF STUDY		EXPERIMENTAL DESIGN		
	RANDOM	INTACT	QUASI-EXPERIMENTAL	EXPERIMENTAL	BLOCKING	FACTORIAL	COVARIANCE
UPPER QUARTILE	.31	.33	.34	.30	1.50	.29	.29
MEDIAN	-.07	.04	.12	-.05	.60	-.01	.01
LOWER QUARTILE	-.30	-.09	-.07	-.29	-.02	-.30	-.19
$\overline{ES}$	.24	.06	.11	.23	1.22	.16	-.01
Ses	1.38	.41	.39	1.33	1.86	1.20	.40
N (E.S.)	132	58	90	122	18	119	72
NO. STUDIES	12	9	12	12	6	9	8

CONTINUED



	TYPE OF MEASUREMENT		REACTIVITY			RELIABILITY			
	PUBLISHED	UNPUBLISHED	LOW	MEDIUM	HIGH	40-69	70-79	80-89	90-95
UPPER QUARTILE	.37	.29	.27	.49	.57	.24	.33	.06	.04
MEDIAN	.13	-.08	-.02	.10	.38	.07	.11	-.08	-.08
LOWER QUARTILE	-.02	-.36	-.25	-.11	.16	-.23	-.14	-.49	-.24
$\overline{ES}$	.39	.09	.10	.44	.35	.02	.08	-.15	-.08
Ses	1.09	1.12	.91	.97	.27	.46	.39	.38	.28
F(E.S.)	68	144	178	18	15	31	67	32	26
N STUDIES	15	17	21	4	4	6	12	9	5

CONTINUED



UPPER QUARTILE	.29	.29	.34	.53
MEDIAN	-.01	-.02	-.02	-.11
LOWER QUARTILE	-.21	-.23	-.41	-.41
$\overline{ES}$	.11	.11	.31	.33
Ses	.72	.76	1.56	1.66
(E.S.)	135	118	61	33
N[illegible]TUDIES	14	13	9	5

There was little difference in resultant effect size between those studies with random assignment of subjects and those which utilized intact groups. Those with intact groups did, however, have less variation in values. In addition, little difference was found when comparing the quasi-experimental and experimental studies. Differences were observed between those studies using a simple blocking design and those using a factorial or covariant design which may suggest that different conclusions may be drawn depending upon the experimental design. In this case, however, further analysis indicated that there was insufficient data for a separate analysis.

A comparison between the effect sizes for those studies which utilized nationally published as opposed to ad hoc published outcome measures showed little difference. Differences were detected, however, between those studies using highly reactive measures and those having moderate to low reactivity. A separate analysis was conducted for those studies having a low outcome reactivity even though there were notably fewer studies in the medium and high categories leading to the possibility that they may not have of themselves influenced the results. In addition, the effect sizes for those studies using an outcome measurement with a reliability $\leq .79$ were dissimilar from those of studies having instrumentation with a reliability of $\geq .80$. Insufficient data, however, was available for an analysis with studies having outcome reliabilities $\geq .80$. Finally, little variation was found in the effect sizes across studies depending upon how the effect size was calculated.

The frequencies analysis and crosstabulation of study characteristics resulted in the selection of 30 variables which provided adequate data for comparative analysis. The results of this analysis, with 21 variables meeting the criteria for inclusion, are shown in Figure 1. If different results were obtained by the selection of comparisons having low reactivity only the results from the low reactive comparison are provided.

Outcome measures were categorized along the following three dimensions: (1) the intent of the assessment acquisition, i.e. the use of identical information as compared to dissimilar information, (2) the intent of assessment transfer, and (3) the assessment domain of orientation, i.e., knowledge, application, process, or problem-solving.

TABLE 2

CORRELATION COEFICIENTS FOR STUDY CHARACTERISTICS OF INDUCTIVE VS. DEDUCTIVE RESEARCH

	SC03	SC16	CC02	TD01	TD02	EX06	EX07	EX11	EX30	EX32	EX33
SC02	.86										
SC10			.56		.65	.63					
SC15		.97									
CC02					.57	.59	.51				
DC01				.59	-.53	-.54					
DC03						.61	-.55				
DC05				-.64	.53	.53	-.53				
TD02						.65	-.68				
EX06							-.87	-.68	-.53		
EX07								.72			
EX08								.65			
EX11									.83		
EX18											-.58
EX31										.80	

META - ANALYSIS DATA FILE VARIABLES AND DEFINITIONS

SC02 - Ability level (IQ)

SC03 - Homogenity of IQ: (1) Homogeneous (2) Heterogeneous

SC10 - SES: (1) Low (2) Low & Medium (3) Medium (4) Medium and High (3) High

SC15 - Class size (no. of students): Experimental

SC16 - Class size (no. of students): Control

CC02 - Community type: (1) Urban (2) Rural (3) Suburban (4) Mixed

DC01 - Assignment of Subject Treatments: (1) Random (2) Matched (3) Intact Groups (4) Self Select

DC03 - Rated Internal Validity (see conventions): (1) Low (2) Medium (3) High

DC05 - Type of Study: (1) Correlational (2) Quasi-Experimental (Descriptive) (3) Experimental (4) Pre-Experimental (One Group Pre-Post)

TD01 - Number of weeks

TD02 - Number of sessions

EX06 - Inductive vs Deductive: (1) Inductive (Discovery) (2) Deductive (Expository)

EX07 - Guidance: (1) Structured (2) Free exploration (3) Guided Exploration

EX08 - Level of Access: (1) Remore demonstration (2) Individual manipulation

EX11 - Levels of Inquiry (see Shulman & Tamir, 1973): (1) None (2) Low (3) Medium (4) High

EX14 - Mathemagenic Behaviors (see Rothkopf, 1970): (1) Used (2) Translation (3) Segmentation (4) Processing

EX30 - Degree of Generality: (1) Items (2) Categories (3) Systematic Patterns

EX31 - Type: (1) Progressive Differentiation (2) Developmental Level of Cognitive Functioning (3) Hierarchical (4) Random (5) Learning cycle (i.e., SCIS)

EX32 - Sequencing Unit. (1) Single lesson (2) Instructional unit (3) Instructional Term (4) Instructional Program

EX33 - Content-orientation (see Klopfer, 1971): (1) General Science (2) Biological Sciences - 10-24 (3) Chemistry - 26-35 (4) Physics - 41-48 (5) Earth Sciences - 56-60 (6) Biochemistry

There is an indication that the inductive approach does not help students when the evaluation criteria is the acquisition of identical information. For this criteria the students taught with an inductive approach were -.22 of a standard deviation from the average member of the control group. When the evaluation instrumentation called for the demonstration of a capability with similar concepts the inductive group was only .05 of a standard deviation from the control group. An examination of the remaining outcome characteristics shows that there was little difference between the inductive and deductive groups for transfer, comprehension, or application of concepts, as well as process skills and problem solving.

Intermediate students seemed to perform better within an inductive-oriented setting with a .18 standard deviation difference. The average intermediate student would do better than 57% of those within a deductive approach. Moreover, 75% of those taught with the inductive approach would perform better than 52% of those in the deductive group. There was little difference at the junior high level and the average subject in high school who was exposed to this approach actually performed not better than 44% of those in the deductive-oriented approach.

There seems to be little differences in approach depending upon ability level, homogenity of ability or gender. Those having an IQ of 93 - 107 performed as well as 47% of those in the deductive group. Subjects in heterogeneous groups accomplished nearly as much as those in the control group while those in homogeneous groups performed as well as 42% of those in the deductive group. Moreover, distinctions based upon seriation ability indicate little difference in performance. There are differences, however, with respect to class size. Those in classes of 17-26 performed better when experiencing an inductive approach, with the average subject performing better than 56% of those in the deductive group. As class size increased performance in comparison to the deductive group decreased.

The community context in which studies were conducted had little relationship to student performance. Variation in context never resulted in more than a difference of .08 standard deviations. The duration of the study also seemed to have little effect upon the accomplishments of the inductive group. In each time span the average subject in the treatment group was within .1 standard deviation of those in the deductive group.

Variations in a number of the treatment characteristics seemed to make little difference in the performance of the inductive group compared to those having a deductive approach. These characteristics included the content-orientation, whether the materials included text and manipulative or manipulative only, or whether teacher interaction was direct or indirect. In addition, it did not seem to make a difference whether the learning experiences were intended to develop an understanding of categories or systematic patterns.

The level of inquiry (Shulman and Tamir, 1973) also did not seem to affect the performance. It should be noted, however, that more than 60% of the studies involved students in a medium level of inquiry. The potential for a multivariate analysis was explored using the level of inquiry of the learning experiences treatment characteristic as an independent variable. It was not pursued, however, due to the limited number of studies in which the level of inquiry was codable.

Several other characteristics seem to affect the results including the level of guidance, and the kind of activities. Subjects experiencing inductive learning through guided exploration performed .2 of a standard deviation better than those with an inductive approach, having a more structured environment. While the average student in the treatment group performed better than 54% of those in the deductive-oriented group, those in a more structured atmosphere could only perform on the average better than 46% of those in the control group. Where the intent of the learning experiences was to have students work with categories and organize knowledge into new patterns they performed on the average better than 52% of the control group, whereas those who were required to make distinctions performed only better than 49% of those having a deductive approach. In fact, the results indicate that 75% of those in the inductive approach did better than 48% of those in the deductive-oriented group.

Variations in several other characteristics also seem to affect the results. These features include the instructional sequencing used and the mode of communicating knowledge. Subjects experiencing an inductive curriculum organized with the progressive differentiation of concepts performed .19 standard deviation better than those having a hierarchical inductive curriculum; this was actually .24 standard deviation when including outcome measurements having medium and high reactivity. Moreover, where the framework for the sequencing of concepts was the instructional program, students performed better than when the sequencing unit was the instructional unit. Seventy-five percent of those using a program sequence performed better than 50% of the deductive group and better than nearly 75% of those having the unit as the sequencing feature. Where the emphasized mode of communicating knowledge was through discussion the average subject performed better than 56% of the deductive group, while there was no difference between inductive and deductive when both discussion and lecture were utilized.

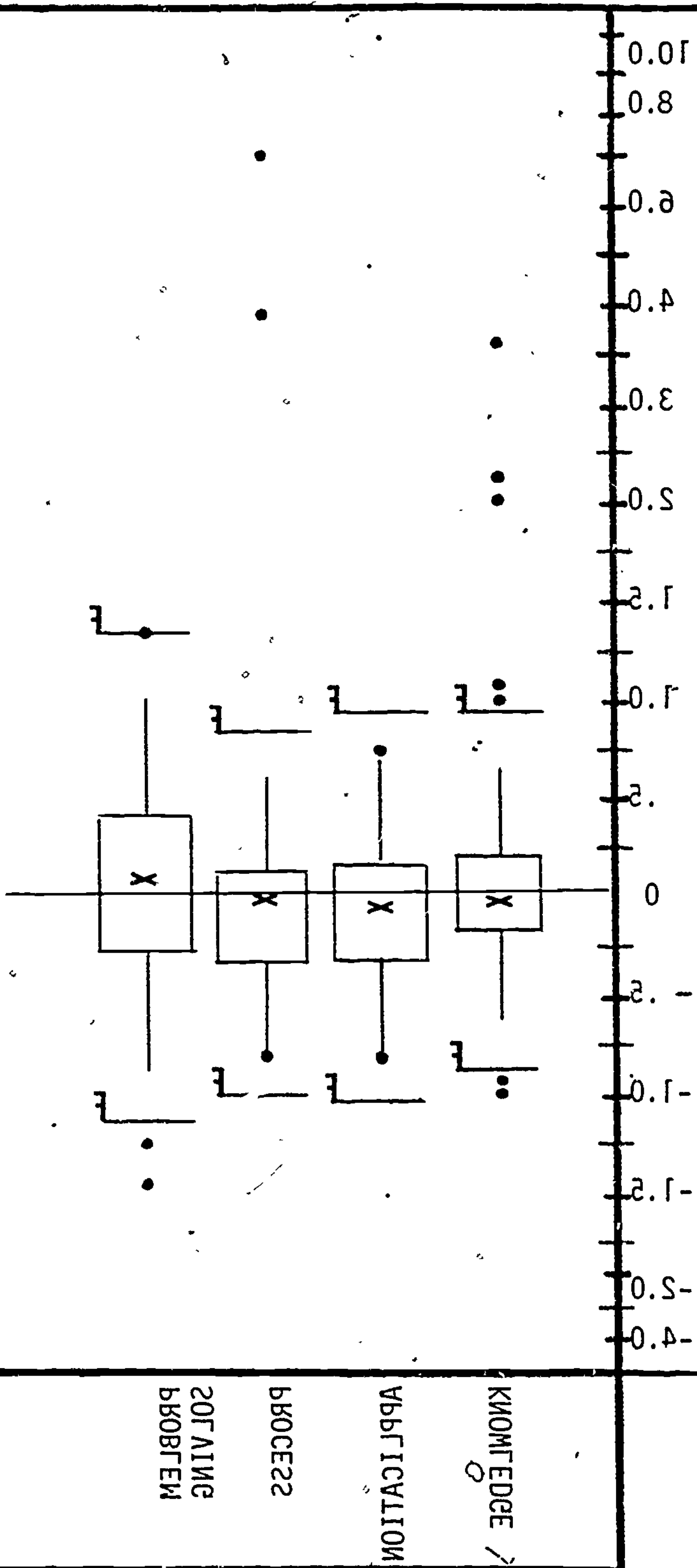
Implications. The results of this analysis comparing the inductive and deductive approaches provide a framework for conjectures as well as directions for future research. Conjectures which seem justified include the apparent positive effect the inductive approach has at the intermediate level. Moreover, this approach seems to be more useful in those situations where high levels of thought, learning experiences, and outcome demands are placed upon the subjects. In addition, the inductive approach appears to function better when the curricular organization is formulated across units to involve the complete program.

It was realized early in the analysis that more research concerning the level of inquiry needs to be conducted. Many studies indicate their concern with inquiry but few address the level of inquiry involved. The difference found between the effect of inquiry experiences upon comprehension and process skills outcome needs to be further explored. This might include the collection of qualitative data concerning treatments in an effort to explore their nature and characteristics. This should provide an insight into the difference in effect.

Several suggestions for future research can be ascertained from the results of this meta-analysis. It would be useful to conduct studies having a range of inquiry levels and utilizing variations in characteristics of manipulative involvement, curricular organization, and approach to the communication of knowledge.

Those conducting research in this area should consider the reactivity and reliability of outcome instruments. Where more reactive instruments must be used, researchers should increase the collection and description of

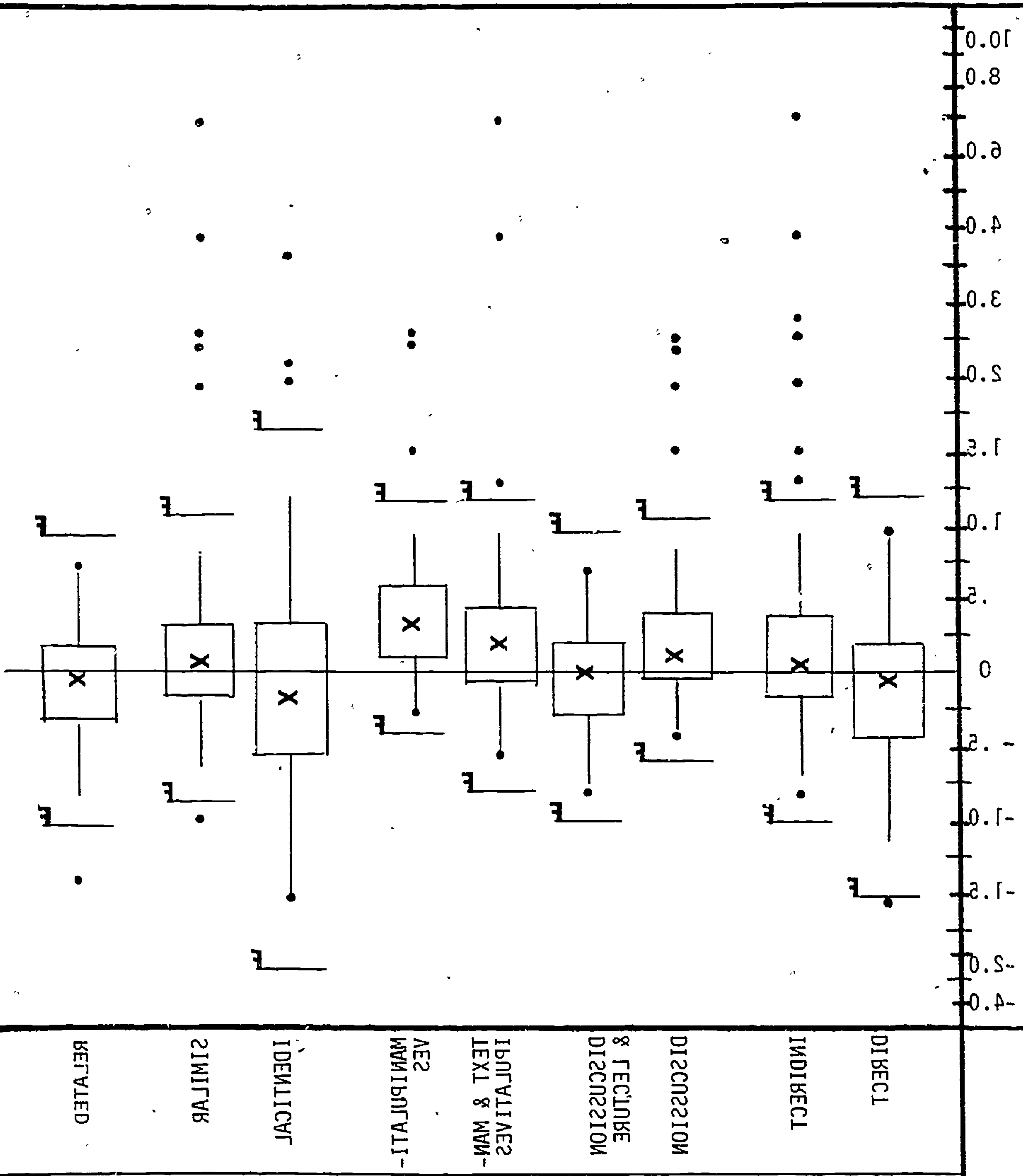
CONTINUED



ASSESSMENT DOMAIN OF ORIENTATION

	KNOWLEDGE	APPLICATION	PROCESS	PROBLEM SOLVING
UPPER QUARTILE	.22	.16	.14	.36
MEDIAN	-.06	-.12	-.02	.03
LOWER QUARTILE	-.24	-.35	-.34	-.27
ES	.02	-.10	.29	-.01
Ses	.67	.36	1.57	.57
F(E.S.)	84	22	28	34
NO. of STUDIES	18	4	8	8

CONTINUED



	TEACHER INTERACTION		MODE OF COMMUNICATING KNOWLEDGE		CHARACTERISTICS OF MATERIALS		INTENT OF ASSESSMENT AQUISITION		INTENT OF ASSESSMENT TRANSFER
	DIRECT	INDIRECT	DISCUSSION	DISCUSSION & LECTURE	TEXT & MANIPULATIVES	MANIPULATIVES	IDENTICAL	SIMILAR	RELATED
UPPER QUARTILE	.21	.39	.41	.23	.45	.57	.30	.29	.17
MEDIAN	-.06	.02	.14	.00	.23	.32	-.22	.05	-.02
LOWER QUARTILE	-.48	-.17	-.01	-.27	-.06	.15	-.63	-.20	-.35
$\overline{ES}$	-.15	.24	.31	-.04	.57	.54	.09	.21	-.11
Ses	.53	1.00	.64	.42	1.48	.73	1.11	1.07	.48
F(E.S.)	63	95	49	35	29	23	25	78	47
NO. of STUDIES	5	13	7	4	6	5	8	17	

qualitative data. The increased use of qualitative data may be helpful in further exploring the difference in effect sizes between instructional sequences based upon progressive differentiation and those with hierarchical sequences. Data collection expanded to include qualitative data is more practical through the use of probit transformations in the calculation of effect sizes. In addition, any increase in qualitative description would assist in the coding of study characteristics.

Experimental research in this area should include less structured learning environments for the inductive approach as well as more rural and suburban contexts for the studies. Future quasi-experimental studies need better documentation of research features as well as more deductive treatments.

Advance Organizers

Advance organizers (Ausubel, 1960, 1963) were proposed to improve "meaningful verbal learning" through the association of what is to be learned with the learner's current conceptual framework. The data base on this topic included 147 effect sizes from 16 studies where the treatment or control group was coded as using an advance organizer. The data base, most of which was not included in the meta-analysis reported by Kozlow and White (1980), is limited mainly to dissertations and science education. Due to the limited data base, a multivariate analysis was not possible.

Table 3 shows the mean effect size for advance organizers vs a control group on several outcome measures. Effect sizes in favor of the advance organizer group are labelled positive. The composite of the two outcome categories has an average effect size of .24. In the aggregate, advance organizers have an advantage of about one quarter of a standard deviation over a control.

Numerous variables were not analyzed due to the lack of variation in study characteristics across studies within this research topic. These features, however, did provide added detail concerning the nature of treatments. There was very little variation in the features of advance organizer research across studies. Most experimental arrangements did not

TABLE 3

Mean Effect Sizes for Advance Organizers vs. Control on Two Outcome Measures and a Composite

	$\bar{\Delta}$	S	n
Knowledge	.09	.59	17
Application	.77	.47	5
Composite	.24	.63	22

use any grouping, were disciplinary in scope, and were interested in student comprehension as an outcome. Only 11.6% of the treatment comparisons used application as an outcome variable. Little variation was found in the type of measurement with a sizable 93.2% of the treatment comparisons using an ad hoc published instrument for outcome measurement. The method of measurement was also seldom different with 88.4% being multiple choice. While there was more variation found in the content-orientation of advance organizer research it seemed apparent that the physical sciences were the most popular at 55.3%.

A correlation analysis of the variables within this research topic resulted in some statistically significant correlations (where $p < .01$). Those with $r \geq .49$ and having important methodological and educational implications are reported in Table 4.

Table 4

TABLE 4

CORRELATION COEFICIENTS FOR STUDY CHARACTERISTICS OF ADVANCE ORGANIZER RESEARCH

	ID05	SC01	SC05	SC14	SC15	CC02	DC01	TD01	TD02	EX25
SC16					.98					
CC02				.52						
DC02						-.58	-.71			
DC03			-.53							
DC05						-.65	-.88			
DC06		-.64								
TD02	-.66							-.56		
EX25	-.64				.58			.68	-.68	
EX54						.69		.64	-.66	.64

META-ANALYSIS DATA FILE VARIABLES AND DEFINITIONS

ID05 - Year of study

SC01 - Modal grade

SC05 - Gender (% female)

SC14 - Special grouping: (1) Not grouped (2) Low track (3) Medium track (4) High track (5) Voluntary

SC15 - Class size (no. of students): Experimental

SC16 - Class size (no. of students): Control

CC02 - Community type: (1) Urban (2) Rural (3) Suburban (4) Mixed

DC01 - Assignment of Ss to Treatments: (1) Random (2) Matched (3) Intact groups (4) Self-select

DC02 - Assignment of teachers to treatments: (1) Random (2) Non-random (3) Self-Select (4) Crossed (5) Matched (6) Investigator

DC03 - Rated Internal Validity (see conventions): (1) Low (2) Medium (3) High

DC05 - Type of study: (1) Correlational (2) Quasi-Experimental (Descriptive) (3) Experimental (4) Pre-Experimental (One Group Pre/Post)

DC06 - Experimental Design: (1) Blocking (10) Factorial (30) Covariance (31) Covariance Blocking (32) Covariance Factorial (33) Covariance Blocking & Factorial

TD01 - Number of weeks

TD02 - Number of sessions

EX25 - Scope of Content: (1) Disciplinary (2) Integrated (3) Multi-Disciplinary (4) Interdisciplinary

EX54 - Text: (1) Text only (2) Text and manipulatives (3) Manipulatives only

The following findings are evident in Table 4.

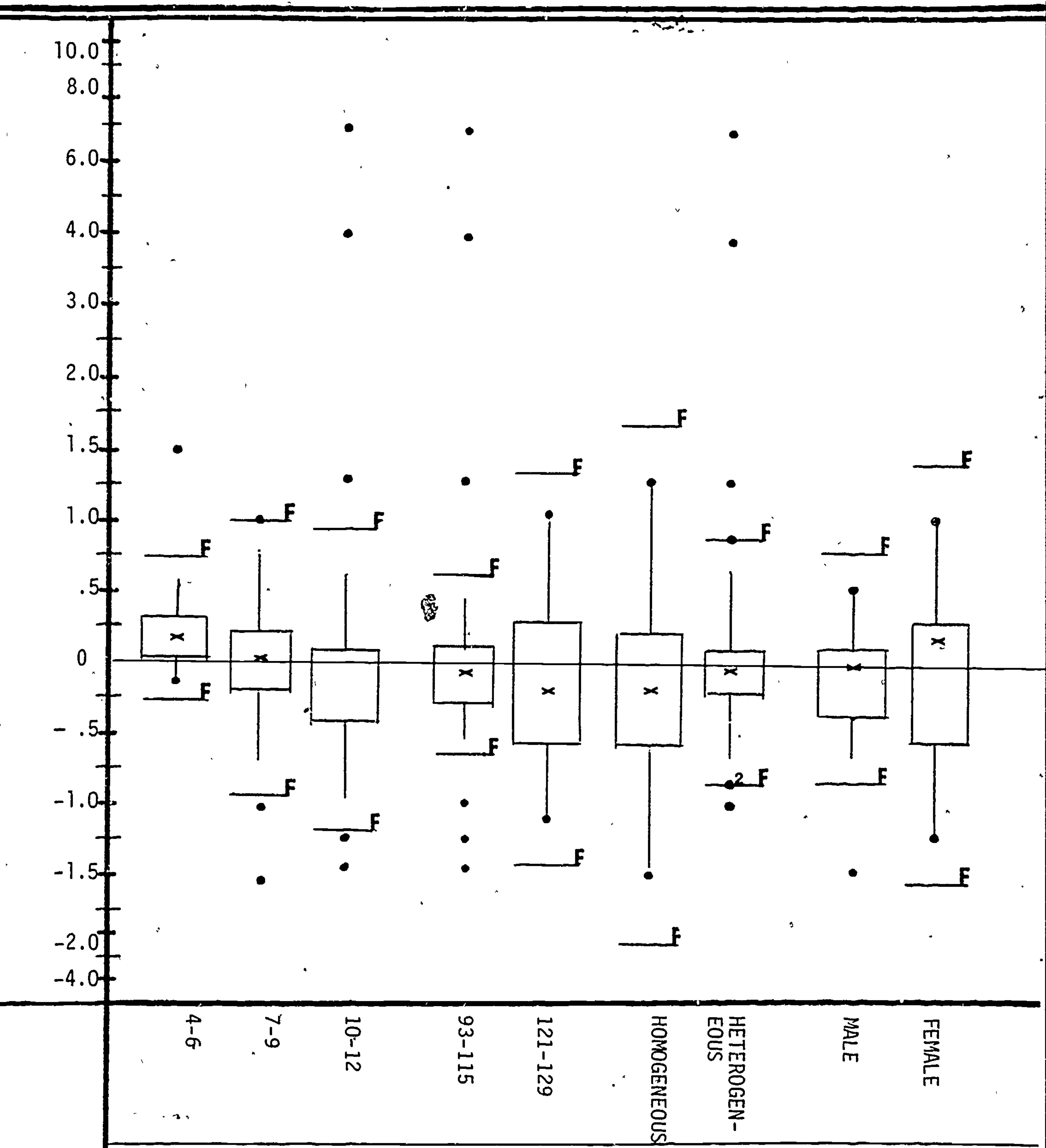
1. The grouping of students was generally practiced more in the suburban environment than the urban or rural.
2. The more recent studies used fewer sessions than earlier studies.
3. In relation to design characteristics those studies having random assignment of subjects did not also randomly assign teachers. If teachers were randomly assigned the subjects were generally from intact classes.
4. Those studies with female participants generally had low internal validity while those with male participants were generally higher.
5. Experimental studies were usually conducted in the urban and rural environment.
6. Studies which utilized designs intended to control for confounding variables were usually conducted at lower grade levels than those which used simpler designs.
7. Recent studies were more inclined to utilize a more multi- or inter-disciplinary approach than earlier studies.
8. It was not surprising that those studies conducted more recently involved the use of manipulatives more than earlier studies.
9. Sample sizes for the treatment groups and the control groups were nearly equal.
10. Manipulatives were generally included as an aspect of studies in suburban environments whereas treatments were more textual-oriented in suburban settings.
11. In relation to the duration of the study those using a multi- or inter-disciplinary organization tended to be longer in duration, however, there were fewer sessions. A similar inverse relationship

existed in relation to the characteristic of instructional materials. While those involving the use of manipulatives tended to be longer duration studies they used fewer sessions per study.

An examination of design characteristics relative to study validity indicated several differences which are important for future methodological decisions. These results are provided in Figure 2. As with the previous research topic these comparisons were made to discern if a separate analysis was required for those studies with a stronger design.

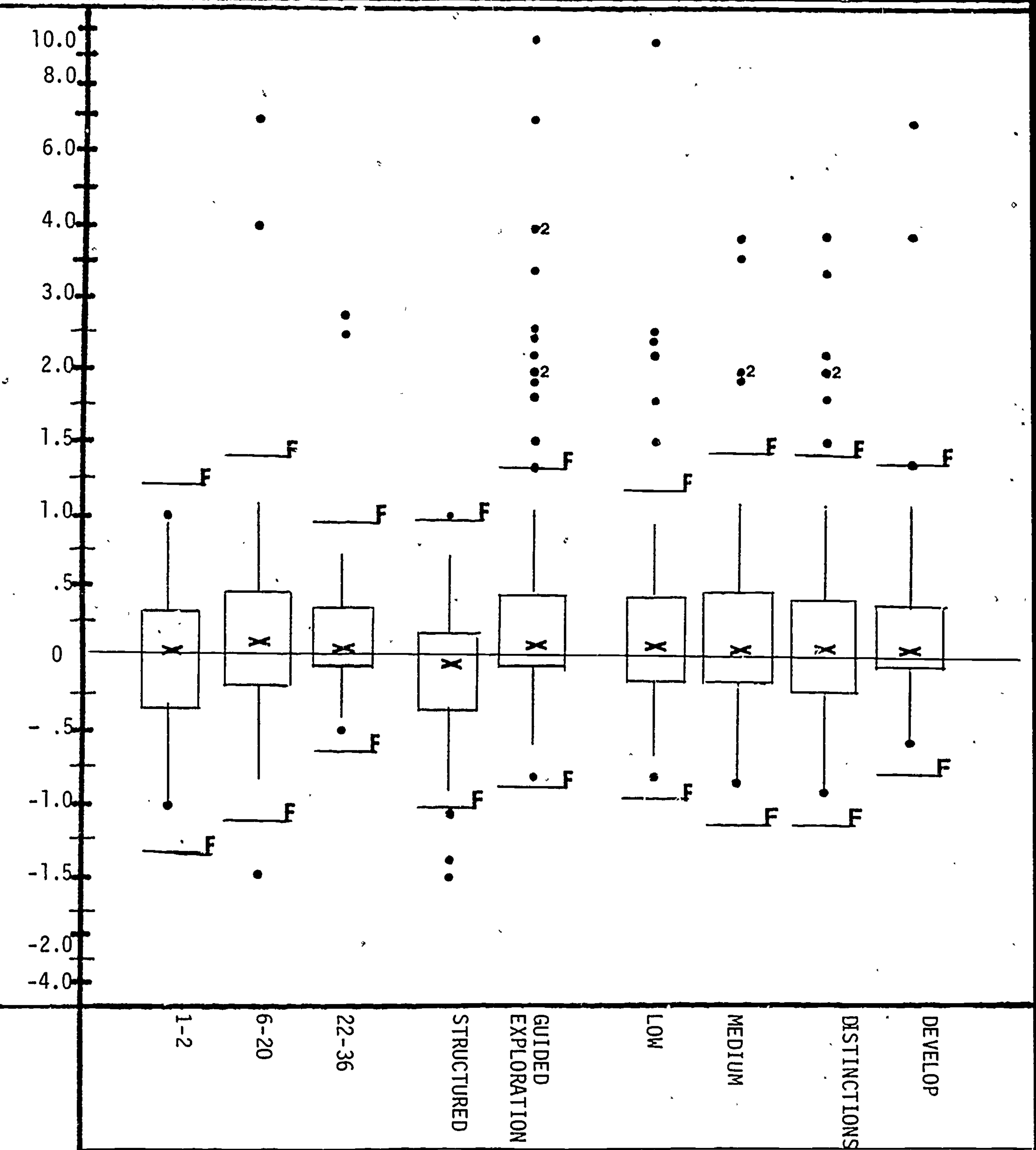
Figure 2

FIGURE 2 BOX-AND-WHISKER PLOTS FOR STUDENT, CONTEXT, TREATMENT, AND OUTCOME CHARACTERISTICS



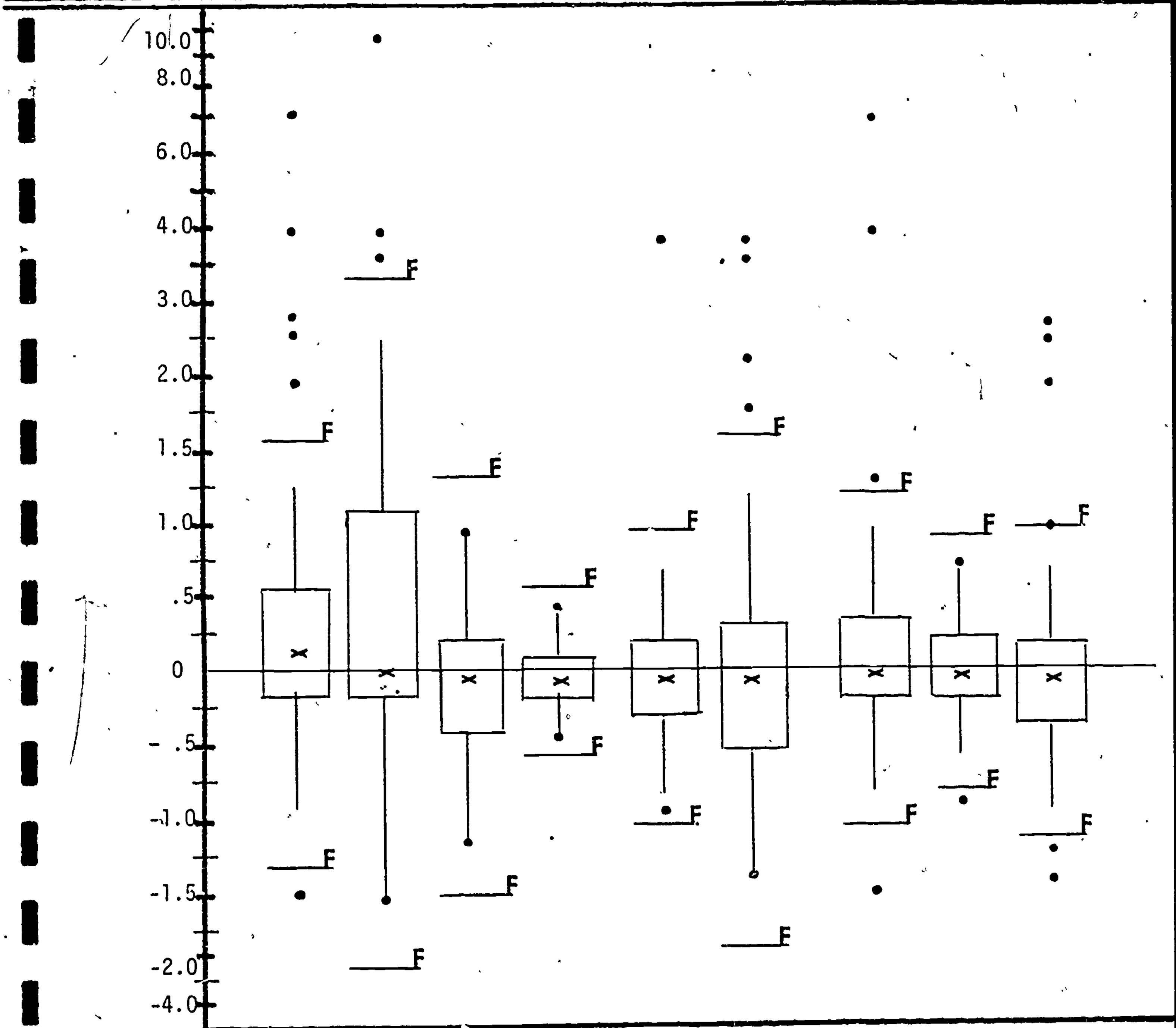
	GRADE LEVEL (SELECTED FOR LOW REACTIVITY)			ABILITY LEVEL: IQ (SELECTED FOR LOW REACTIVITY		HOMOGENITY OF IQ (SELECTED FOR LOW REACTIVITY)		GENDER (SELECTED FOR LOW REACTIVITY)	
	4-6	7-9	10-12	93-115	121-129	HOMOGENEOUS	HETEROGENEOUS	MALE	FEMALE
JPPER QUARTILE	.31	.24	.13	.15	.27	.24	.16	.16	.32
MEDIAN	.18	.02	-.15	-.05	-.21	-.19	-.02	.00	-.21
OWER QUARTILE	.03	-.22	-.41	-.27	-.63	-.63	-.19	-.35	-.56
$\overline{ES}$	.24	-.03	.14	.09	-.19	-.22	.20	-.14	-.12
Ses	.37	.42	1.52	1.13	.56	.58	1.24	.48	.50
(E.S.)	16	1.01	34	63	19	34	48	27	27
NO [illegible]TUDIES	6	7	6	7	4	4	7	5	5

CONTINUED



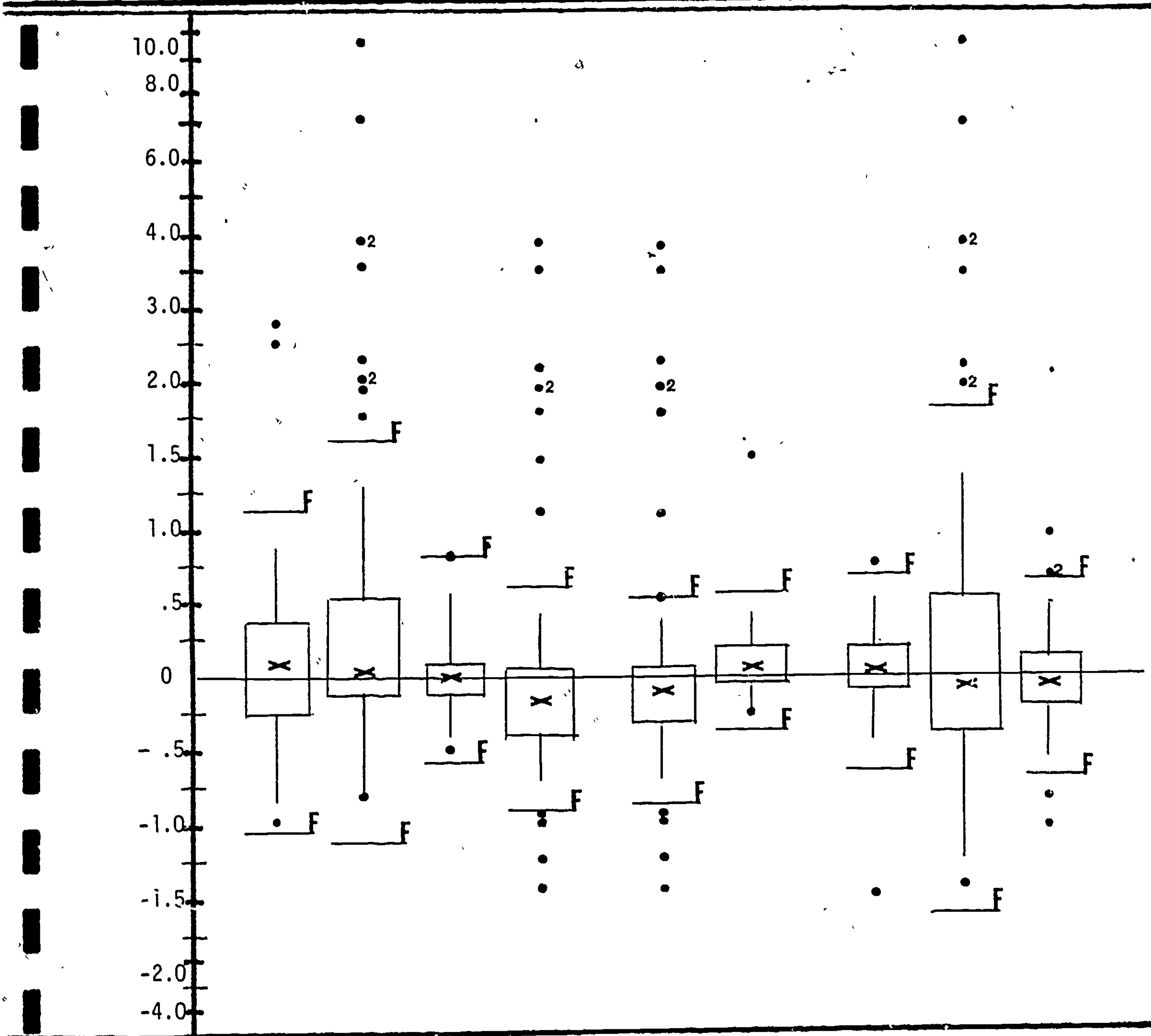
	STUDY DURATION (NO. OF WEEKS)			GUIDANCE		LEVEL OF INQUIRY		KINDS OF ACTIVITIES (SELECTED FOR LOW REACTIVITY)	
	1-2	6-20	22-36	STRUCTURED	GUIDED EXPLORATION	LOW	MEDIUM	DISTINCTIONS	DEVELOP
UPPER QUARTILE	.27	.43	.33	.16	.45	.39	.46	.35	.36
MEDIAN	.01	.11	.03	-.09	.10	.07	.04	-.03	.04
LOWER QUARTILE	.38	-.23	-.07	-.39	-.11	-.17	-.20	-.25	-.07
$\overline{ES}$	-.05	.28	.22	-.14	.43	.29	.41	.32	.51
Ses	.43	1.29	.69	.48	1.37	1.25	1.14	1.06	1.49
F(E.S.)	48	46	30	79	1.22	84	31	44	29
N STUDIES	4	5	6	8	15	12	5	5	6

CONTINUED



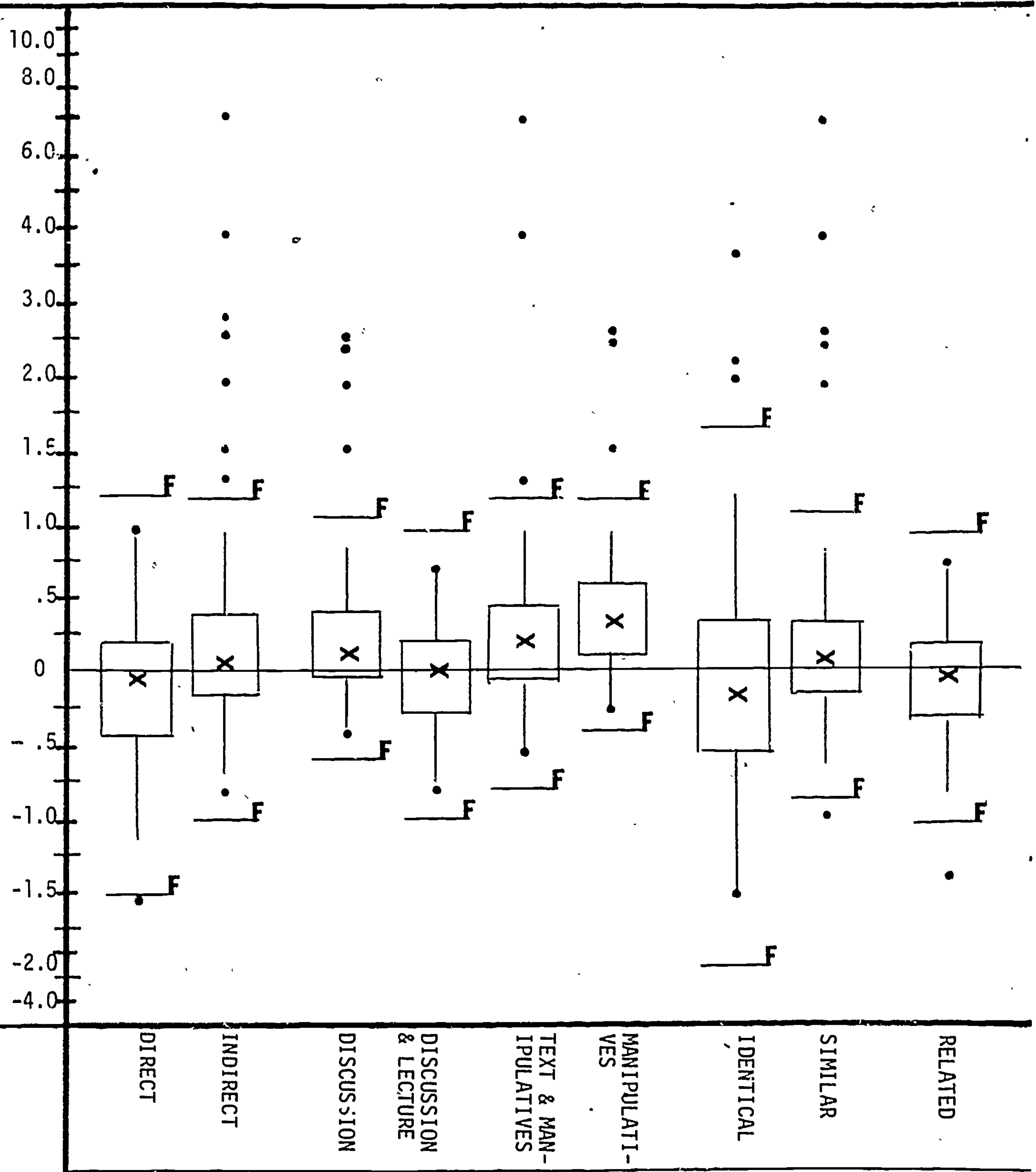
	EXP 17-26 CON 17-26	EXP 28-32 CON 28-33	EXP 39-58 CON 39-63	EXP 85-172 CON 92-216	STAGE III	STAGE I	URBAN	RURAL	SUBURBAN
	CLASS SIZE				SERIATION ABILITY		COMMUNITY TYPE		
UPPER QUARTILE	.52	1.15	.23	.10	.21	.33	.36	.24	.17
MEDIAN	.16	-.01	-.10	-.08	-.11	-.15	-.05	.03	-.10
LOWER QUARTILE	-.21	-.20	-.49	-.18	-.30	-.56	-.23	-.19	-.39
$\overline{ES}$	.26	.74	-.11	-.05	.22	.18	.28	.01	-.05
Ses	1.10	2.04	.42	.23	1.83	1.07	1.29	.40	.65
F(E.S.)	83	33	12	26	38	33	44	39	84
NO. OF STUDIES	8	5	5	5	5	5	7	4	7

CONTINUED



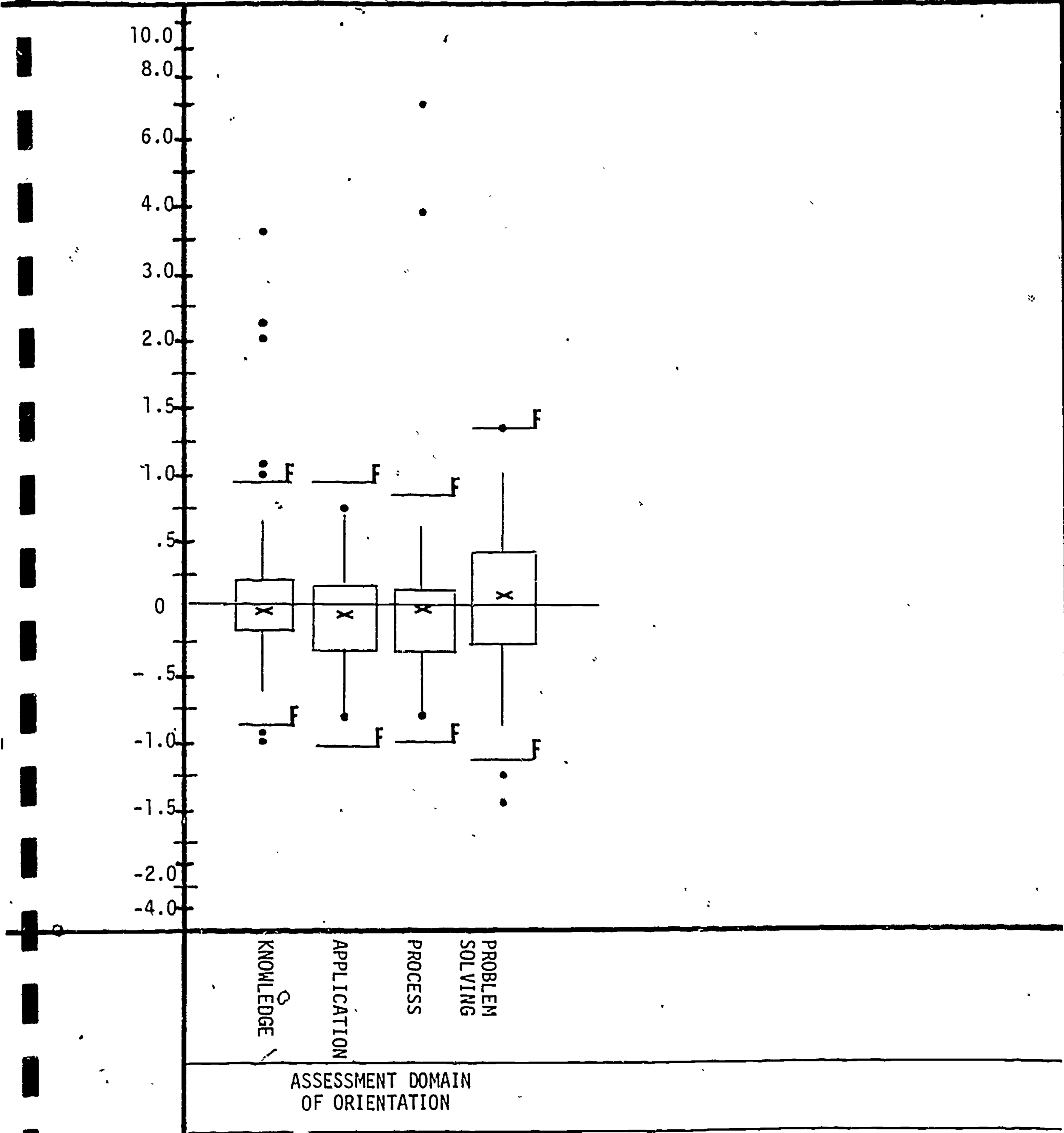
	CATEGORIES	SYSTEMATIC PATTERNS	PROGRESSIVE DIFFEREN-TIATION	HIERARCHIAL	INSTRUCTION-AL UNIT	INSTRUCTION-AL PROGRAM	BIOLOGY	CHEMISTRY	PHYSICS
	DEGREE OF GENERALITY		INSTRUCTIONAL SEQUENCING (LOW REACTIVITY)		INSTRUCTIONAL SEQUENCING UNIT (LOW REACTIVITY)		CONTENT ORIENTATION		
JPPER QUARTILE	.32	.58	.12	.04	.04	.23	.23	.53	.15
MEDIAN	.11	.06	.00	-.19	-.16	.04	.03	-.11	-.05
.O[illegible]ER QUARTILE	-.25	-.12	-.11	-.36	-.33	-.01	-.11	-.36	-.20
$\overline{ES}$	.08	.68	.02	.13	.07	.15	-.01	.50	-.05
Ses	.58	1.75	.22	1.14	1.03	.36	.47	1.93	.37
[illegible](E.S.)	89	65	26	45	54	19	19	58	65
[illegible]STUDIES	7	8	6	6	6	6	4	6	5

CONTINUED



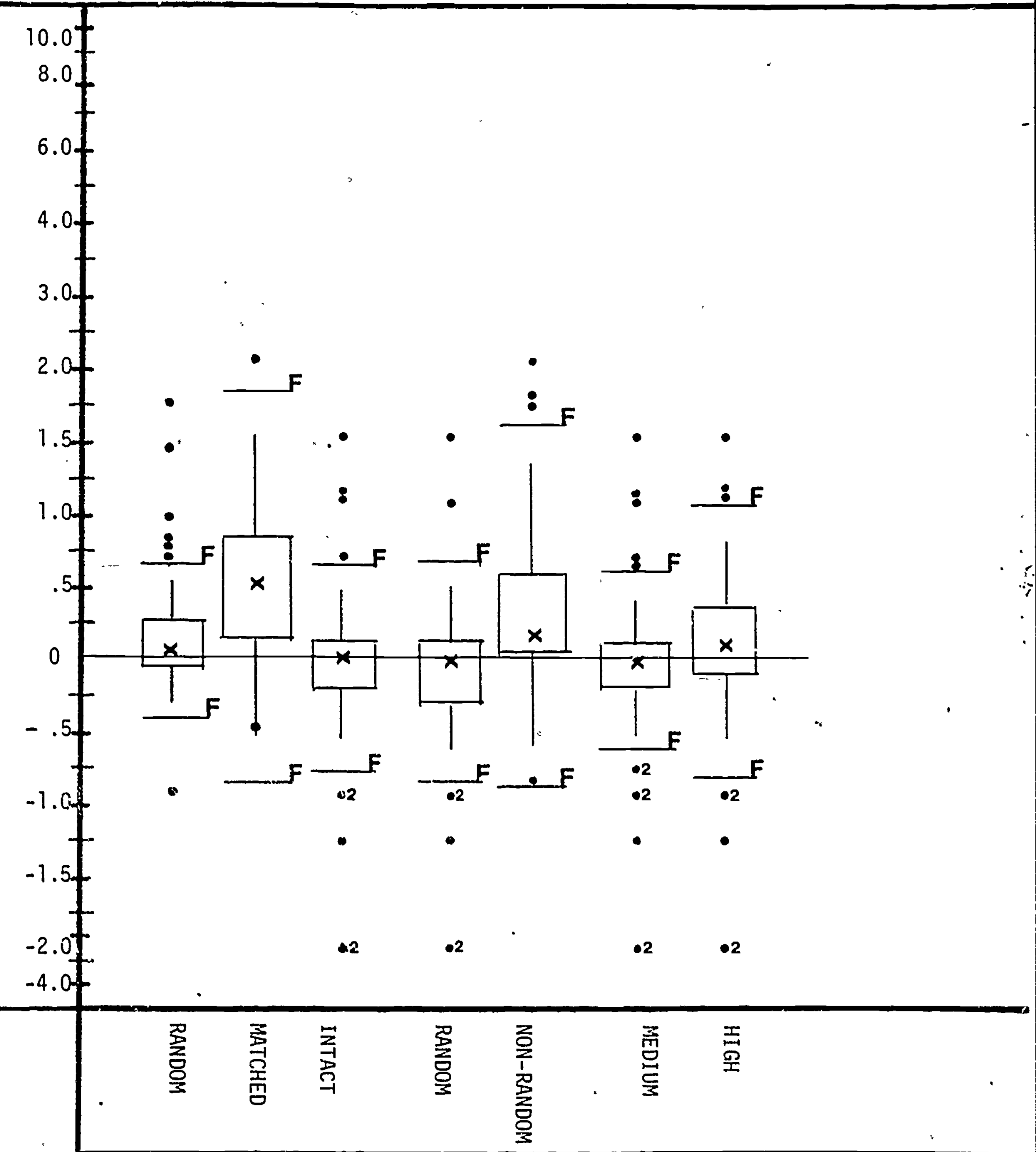
	TEACHER INTERACTION		MODE OF COMMUNICATING KNOWLEDGE		CHARACTERISTICS OF MATERIALS		INTENT OF ASSESSMENT AQUISITION		INTENT OF ASSESSMENT TRANSFER
	DIRECT	INDIRECT	DISCUSSION	DISCUSSION & LECTURE	TEXT & MANIPULATIVES	MANIPULATIVES	IDENTICAL	SIMILAR	RELATED
UPPER QUARTILE	.21	.39	.41	.23	.45	.57	.30	.29	.17
MEDIAN	-.06	.02	.14	.00	.23	.32	-.22	.05	-.02
LOWER QUARTILE	-.48	-.17	-.01	-.27	-.06	.15	-.63	-.20	-.35
$\overline{ES}$	-.15	.24	.31	-.04	.57	.54	.09	.21	-.11
Ses	.53	1.00	.64	.42	1.48	.73	1.11	1.07	.48
F(E.S.)	63	95	49	35	29	23	25	78	47
TUDIES	5	13	7	4	6	5	8	17	7

CONTINUED



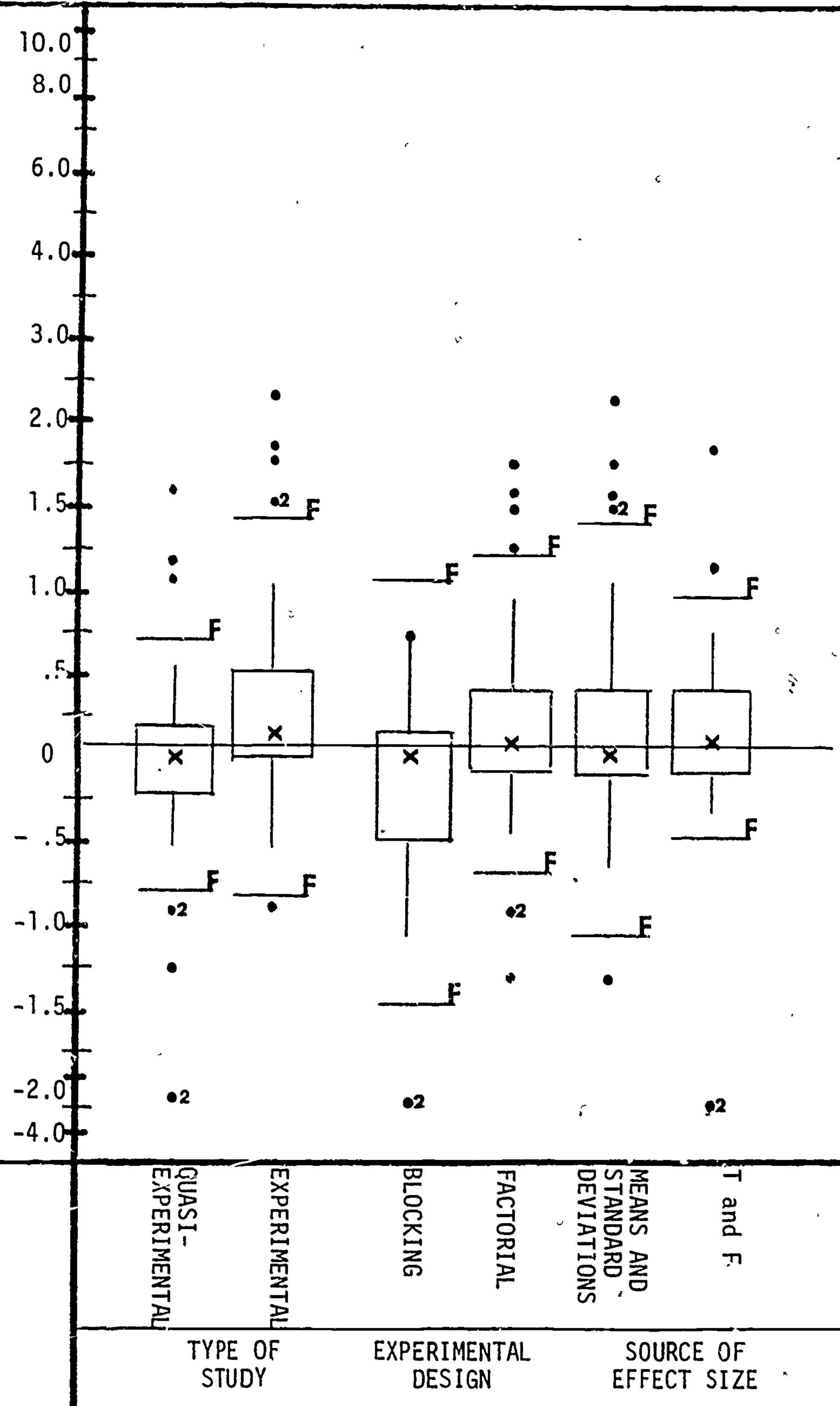
UPPER QUARTILE	.22	.16	.14	.36
MEDIAN	-.06	-.12	-.02	.03
LOWER QUARTILE	-.24	-.35	-.34	-.27
ES	.02	-.10	.29	-.01
Ses	.67	.36	1.57	.57
F(E.S.)	84	22	28	34
N STUDIES	18	4	8	8

FIGURE 3 BOX-AND-WHISKER PLOTS FOR DESIGN CHARACTERISTICS, OUTCOME CHARACTERISTICS, AND SOURCE OF EFFECT SIZE



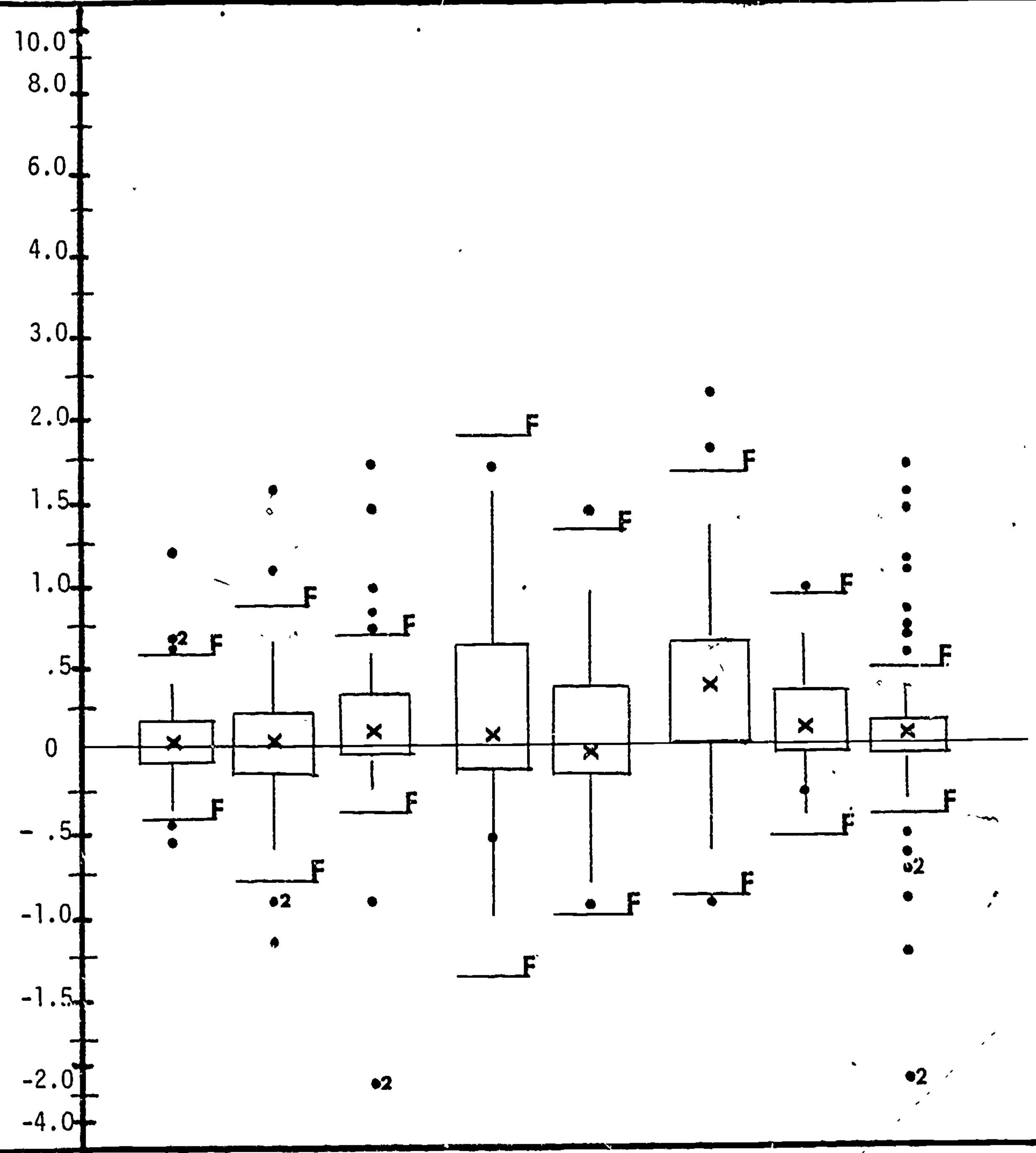
	ASSIGNMENT OF SUBJECTS			ASSIGNMENT OF TEACHERS		INTERNAL VALIDITY	
UPPER QUARTILE	.25	.83	.13	.11	.66	.11	.33
MEDIAN	.09	.50	.00	-.05	.16	-.01	.05
LOWER QUARTILE	-.03	.15	-.24	-.28	.01	-.20	-.13
$\overline{ES}$	.16	.56	-.08	-.13	.35	-.08	.07
Ses	.42	.58	.67	.67	.53	.64	.05
(E.S.)	52	36	52	46	83	57	22
TUDIES	7	4	7	5	9	9	5

CONTINUED



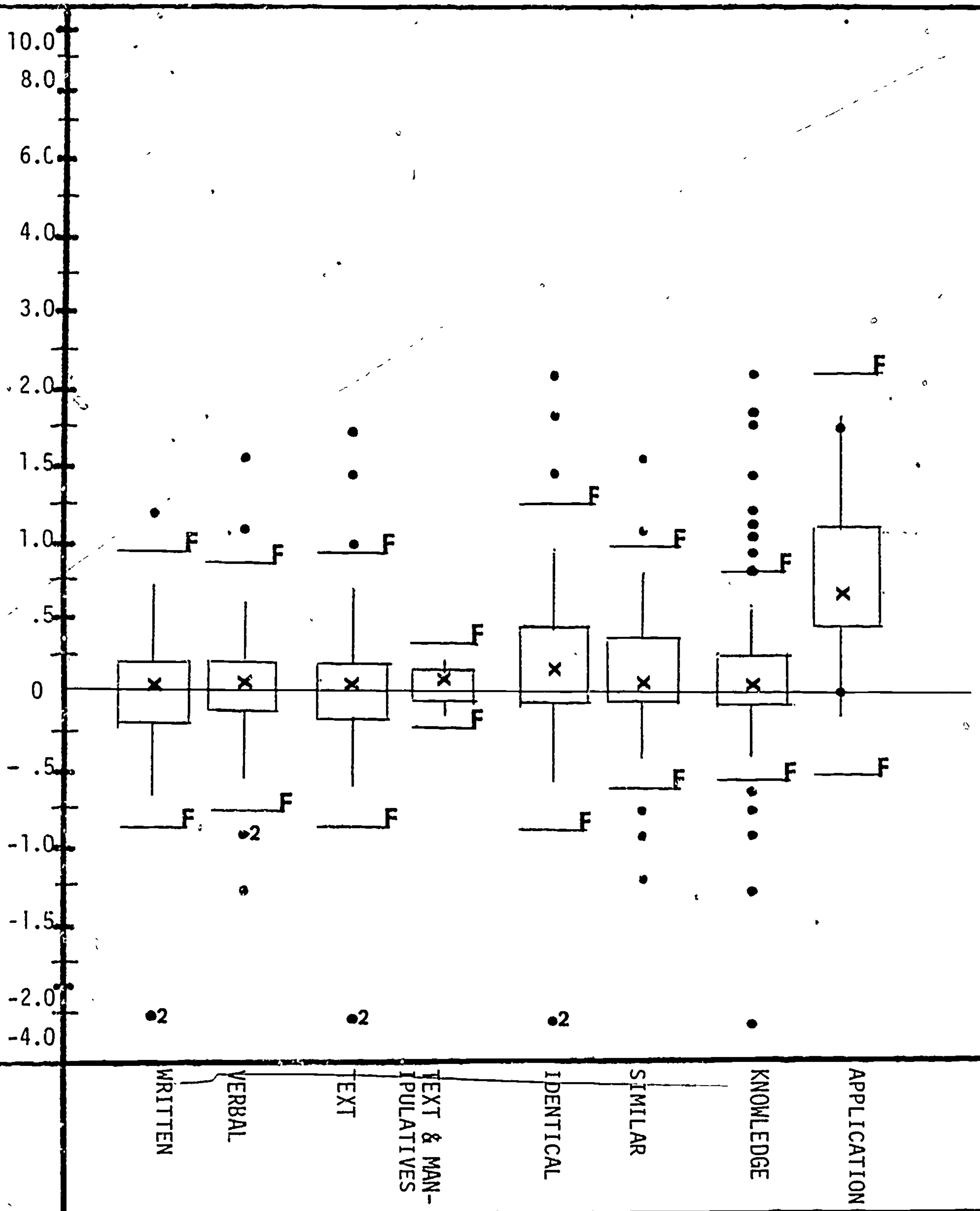
JPPER QUARTILE	.15	.57	.12	.43	.46	.40
MEDIAN	.01	.16	.01	.14	.06	.16
O[illegible]ER QUARTILE	-.22	.00	-.52	-.04	-.14	.05
$\overline{ES}$	-.07	.31	-.28	.20	.17	.16
Ses	.67	.52	.76	.54	.56	.77
(E.S.)	51	89	23	55	107	33
IC TUDIES	07	12	05	08	11	09

FIGURE 4 BOX-AND-WHISKER PLOTS FOR STUDENT CHARACTERISTICS, CONTEXT CHARACTERISTICS, TREATMENT CHARACTERISTICS, AND OUTCOME CHARACTERISTICS



	GRADE LEVEL			SERIATION ABILITY		COMMUNITY TYPE		
	4-6	10-12	13-16	STAGE I	STAGE III	URBAN	RURAL	SUBURBAN
UPPER QUARTILE	.15	.23	.29	.68	.38	.66	.33	.15
MEDIAN	.06	-.03	.11	.05	-.04	.34	.11	.05
LOWER QUARTILE	-.10	-.19	.02	-.17	-.22	.00	-.04	-.09
$\overline{ES}$	.06	.02		.26	.01	.37	.18	.01
Ses	.31	.60		.59	.61	.57	.33	.65
F(E.S.)	46	19	44	15	17	53	15	64
N STUDIES	04	05	06	05	05	05	04	08

CONTINUED



	STYLE OF ADVANCE ORGANIZER		CHARACTERISTICS OF MATERIALS		INTENT OF ASSESSMENT: AQUISITION		ASSESSMENT DOMAIN OF ORIENTATION	
	WRITTEN	VERBAL	TEXT	TEXT & MANIPULATIVES	IDENTICAL	SIMILAR	KNOWLEDGE	APPLICATION
P ER QUARTILE	.24	.23	.24	.10	.46	.38	.25	1.13
MEDIAN	.02	.05	.09	.05	.16	.07	.07	.68
O ER QUARTILE	-.23	-.17	-.22	-.03	-.07	-.03	-.09	.47
$\overline{ES}$	-.09	.02	-.01	.01	.18	.12	.09	.77
Ses	.77	.61	.77	.20	.89	.52	.59	.47
(E.S.)	24	25	35	41	34	50	118	17
UDIES	05	04	04	04	09	11	17	05

A comparison between the effect sizes for those studies which had high internal validity as opposed to medium internal validity showed little difference. In addition, an examination of effect sizes selected for source of effect size data showed little difference. However, there were differences detected between those studies using matched sample and those having random selection or intact groups. The effect sizes for those studies using matching techniques were in general .41 standard deviations higher.

While effect sizes vary somewhat with respect to the selection of teachers, the type of study, and the experimental design, these differences are not very substantial. It was observed that 75% of the effect sizes for studies using non-random assignment of teachers to treatment were greater than 50% of those for studies using random assignment. In the case of type of study 75% of the effect sizes for experimental studies were greater than 50% of those for the quasi-experimental studies. Most of the effect sizes for those studies using a factorial design were greater than those from studies using a block design. These differences suggest that different conclusions may be drawn from the study characteristic analysis depending upon the design characteristics. However, a further analysis of study characteristics indicated that there was insufficient data for a separate analysis based upon the selection of specified design characteristics.

A summary analysis of the advance organizer data base across all study characteristics resulted in finding a median effect size of .09 one hi and hinges of .43 and -.07. Thus, the average student experiencing an advance organizer preinstructional strategy did better than only 54% of the control group. However, the spread between hinges indicates that 75% of those having advance organizers did better than 47% of the control group. The following narrative will describe any differences in effect size based on

ledge

particular study characteristics. This description will show the circumstances under which advance organizers could be expected to provide a more effective educational approach.

The frequencies analysis of study characteristics resulted in the selection of 72 variables which had adequate data for further study. A crosstabulation analysis found 24 variables which had sufficient data for possible comparative analysis when 10 effect sizes in a cross-tab cell was set as the minimum required. The results of this analysis in which 7 variables were found to meet the criteria of 4 or more studies are shown in Figure 4.2

There is an indication that variation in grade level or student seriation ability makes little difference. However, effect sizes did differ depending upon community context. Those studies conducted in a suburban environment had effect sizes which were in general lower than those from studies in urban contexts. The average experimental subject from a suburban context was only greater than 52% of the control group while the average subject in an urban environment scored above 63% of the control group. Moreover, 75% of the effect sizes from the suburban studies were lower than 50% of those from rural studies.

Enough data for analysis was found for only two treatment characteristics; the style of advance organizer and the characteristic of materials. In each case there was little difference between the advance organizer groups and the control groups. The effect sizes for studies using written or verbal advance organizers were similar as were the effect sizes for the studies which used only textbooks or those havint textual as well as manipulative materials.

Only two outcome characteristics had adequate data for analysis. Little difference in effect sizes was found between those studies evaluating student performance on identical information as opposed to similar information. There was, however, a difference between performance on knowledge oriented instruments and application instruments. The performance of the average subject on application was better than 68% of the control group while the performance of the treatment group for comprehension items was better than only 46% of the control group. It should be pointed out however that the application analysis was based upon 17 effect sizes from only five studies while the analysis of the comprehension data was based upon 118 effect sizes from seventeen studies.

Implications. The results of this analysis pertaining to the effect of advance organizers upon student outcomes has provided needed information for establishing directions for future research. In addition, it has influenced the formulation of conjectures concerning the effectiveness of advance organizers.

The data analysis seems to indicate that advance organizers have been advantageous in the urban setting than in rural or suburban contexts. There seems to be little effect depending upon grade level, style of organizer, or characteristics of materials. However, as noted above there has been little variation in treatment or outcome characteristics across studies. It would be useful for future studies to break out of the past advance organizer research pattern and use as yet infrequently applied characteristics.

A further exploration of outcome distinctions is necessary where such features as transfer and application are used. Very little advance organizer research has used application items for the assessment of performance or understanding. Future research should address this question of subject ability to apply what has been taught. Another aspect which should be considered is the extension of study duration and in particular the number of sessions. Moreover, due to the results indicating design characteristic differences in effect it may be appropriate to utilize more experimental and factorial designs utilizing where possible a matching technique.

It would also be worthwhile to compare variations in type of advance organizer across other characteristics to determine any distinct effects based upon type of advance organizer. The data in this study was too limited to pursue that analysis.

CONCLUSIONS AND RECOMMENDATIONS

This meta-analysis may provide a foundation for the continued exploration of learning and teaching in science education. Some research areas should receive more attention, especially the level of inquiry under different curricular treatments. This aspect of curricular variation has not, as evidenced by this meta-analysis, been subjected to any extensive analysis. The duration of experimental studies should be extended and the collection of qualitative datashould be increased for both quasi- and experimental studies. The pursuit of these suggestions should assist the research community in better articulating any distinctions which exist between treatments.

The most limiting aspect of this study has been the lack of descriptive information in studies coded. In addition, many studies were not coded due to insufficient reporting of descriptive or analysis statistics. It is hoped that the coding variables formulated for this study can provide a beginning framework for the design and communication of research characteristics in future studies. The lack of descriptive information, in addition to the limited number of studies coded, resulted in the inability to explore complex interactions and the effect of confounding variables not addressed in individual studies.

The next step should be, in addition to the more complete and thorough description of studies, the continued coding of studies not included in this analysis in order to extend the data base such that further analysis can be undertaken. It is then, with an expanded data base, that the technique of meta-analysis can be used to its fullest potential. In addition, this continued coding could lead to the inclusion of research areas as yet not analyzed. These include behavioral objectives, kinetic structure, mathemagenic behaviors, scope of content and the integrated curriculum hypothesis, and the organization of curriculum.

REFERENCES

Glass, G.V., McGaw, B., and Smith, M.L. Meta-Analysis in Social Research. Beverly Hills, California: Sage Publications, 1981.

Hanna, J.F. "A New Approach to the Formulation and Testing of Learning Models." Synthesis, 1966, 16 344-380.

Hanna, J.R. Explanation, Prediction, Description, and Information Theory. Synthesis. 1969, 20, 308-334.

Popper, K.R. Conjectures and Refutations: The Growth of Scientific Knowledge. London: Routledge and Kegan, 1962.

Yeany, R.H., Jr. Applying and Interpreting the Results of Meta-Analysis Procedures. Paper presented at the meeting of the National Association for Research in Science Teaching, Boston, April 1980.