R E P O R T R E S U M E S

ED 013 227 SE 003 113

COMPILATION OF RESEARCH RESULTS IN ELEMENTARY ARITHMETIC SINCE 1900. FINAL REPORT.

BY- SUYDAM, MARILYN N.
PENNSYLVANIA STATE UNIV., UNIVERSITY PARK

PUB DATE AUG 67

GRANT OEG-1-7-068592-0174
EDRS PRICE MF-$0.50 HC-$3.04 76P.

DESCRIPTORS- *EDUCATIONAL RESEARCH, *ELEMENTARY SCHOOL MATHEMATICS, *RESEARCH, BIBLIOGRAPHIES, CONCEPT FORMATION, DOCTORAL THESES, EDUCATIONAL OBJECTIVES, EVALUATION, INSTRUCTION, LEARNING THEORIES, MATHEMATICS, PENNSYLVANIA STATE UNIVERSITY, UNIVERSITY PARK, PENNSYLVANIA, UNITED STATES OFFICE OF EDUCATION,

RESEARCH RELATED TO THE TEACHING OF ELEMENTARY SCHOOL MATHEMATICS PUBLISHED IN THE UNITED STATES BETWEEN 1900 AND 1965 IS REVIEWED. THE 799 REPORTS INCLUDED IN THE STUDY WERE OBTAINED FROM A REVIEW OF REPORTS IN 50 JOURNALS. REPORTS ARE CATEGORIZED ON THE BASIS OF MATHEMATICAL TOPIC AND TYPE OF STUDY. TOPICS INCLUDED ARE EDUCATIONAL OBJECTIVES AND INSTRUCTIONAL PROCEDURES, TOPICAL PLACEMENT, BASIC CONCEPTS AND METHODS OF TEACHING THEM, MATERIALS, INDIVIDUAL DIFFERENCES, EVALUATION, AND LEARNING THEORY. CRITERIA USED IN EVALUATION INCLUDED (1) PRACTICAL AND THEORETICAL SIGNIFICANCE, (2) CLARITY OF THE PROBLEM DEFINITION, (3) APPROPRIATENESS OF THE DESIGN TO THE RESEARCH QUESTION, (4) CONTROL OF VARIABLES, (5) SAMPLE SELECTION TECHNIQUES, (6) VALIDITY AND RELIABILITY OF MEASURING DEVICES, (7) ANALYTICAL TECHNIQUES, (8) APPROPRIATENESS OF INTERPRETATIONS AND GENERALIZATIONS, AND (9) ADEQUACY OF THE REPORT. SPECIFIC INFORMATION ON STATISTICAL PROCEDURES, VARIABLES CONTROLLED, SAMPLING PROCEDURE AND POPULATION SIZE, TYPE OF TEST, GRADE LEVEL, AND DURATION IS INCLUDED WHERE APPLICABLE FOR EACH REPORT. CONCLUSIONS WHICH APPEAR TO BE CONSISTENT WITH THE DATA IN EACH STUDY ARE REPORTED. APPENDIXES CONTAIN--(1) LISTS OF REVIEWS, TOPICAL SUMMARIES, AND BIBLIOGRAPHIES OF RESEARCH, (2) FREQUENCY TABLES FOR REPORTS ACCORDING TO JOURNAL SOURCE, MATHEMATICAL TOPIC, AND TIME OF PUBLICATION, (3) DETAILS OF OUTLINES AND DEFINITIONS USED IN CLASSIFYING AND RATING STUDIES, AND (4) A COPY OF THE INSTRUMENT USED IN THE EVALUATION OF RESEARCH STUDIES. (AG)

FINAL REPORT

BR- ~~Project No.~~ 6-8592

Grant No. OEG-1-7-068592-0174

PA-24

COMPILATION OF RESEARCH RESULTS IN ELEMENTARY ARITHMETIC SINCE 1900

August 31, 1967

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

COMPILATION OF RESEARCH RESULTS IN ELEMENTARY ARITHMETIC SINCE 1900

Project No. 6-8592
Grant No. OEG-1-7-068592-0174

Marilyn N. Suydam

August 31, 1967

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

The Pennsylvania State University

University Park, Pennsylvania

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

FOREWORD

This final report is a summary of the full project, which is presented elsewhere as a doctoral dissertation:

Suydam, Marilyn N., "An Evaluation of Journal-Published Research Reports on Elementary School Mathematics, 1900-1965," Unpublished doctoral dissertation, The Pennsylvania State University, 1967.

ACKNOWLEDGMENTS

The writer wishes to thank Dr. C. Alan Riedesel who willingly supplied advice and guidance for the project. Appreciation is also expressed to Professors A. Madison Brewer, Clyde G. Corle, Harold E. Mitzel, Jack N. Sparks, John Withall, and Kenneth H. Wodtke, as well as to the many graduate students who helped at various stages.

Carol J. Rockey and Jan E. Richards are thanked for their help in the preparation and typing of this final report.

I. INTRODUCTION

Background

Since the beginning of this century, the field of educational research has become increasingly important. Educational research in this context is considered to include those educational investigations with some degree of scientific procedure and/or control, involving data collection for a specific purpose. This research is generally related to a hypothesis about some aspect of the role of learning in the curriculum.

The realization that the controlled experiment is a feasible technique for exploring many of the problems and issues which face educators caused overwhelming optimism. This hope of a panacea which would resolve all difficulties once and for all led to disillusionment. However, the concept remained that research should help to point the way toward certain decisions, even if many aspects of the educative process are not readily accessible to its tactics.

Never has there been the emphasis on research that has been developing during the past decade and is prevalent today. The need for using the results of research to give direction to the teaching-learning process has been intensified. The development of curriculum reform movements such as that of "modern mathematics" has further accentuated this need within the subject area of elementary school mathematics. Decisions about curriculum innovations must be related to knowledge about curriculum content and methods. A source of such knowledge and a foundation for decisions is research.

Need for the Study

Research of the present and the future must be based on or indicate consideration of research of the past. Implications for needed research or connotations which could lead to creative development are a part of almost all studies which have been done, either overtly or intrinsically.

One of the difficulties which any researcher faces is locating those studies which will be of most use to him. If a researcher is interested in elementary school mathematics, his search of the literature reveals that there has been no single source of information on previous research on the subject. Instead there are various types of lists, no one of which is complete and current. The task of obtaining this information is even more difficult for the teacher who is interested in utilizing the results of research. It is necessary to synthesize the data which have accumulated.

More than a compilation is needed, however. Scrutiny of the literature reveals that there have been many complaints about the deficiencies of educational research. For example, a sample of comments may be considered:

> Unfortunately, we are still using research methods that are inadequate for the solution of the problems we face. . . . Much of the research conducted in education is faulty. Many studies contain flaws that automatically make them null and void from the standpoint of application. These errors cover all aspects of research. . . . (Mouly, 1963, 395)

> Little of any value can be derived from the tons of research that have been conducted, and the majority of the studies are unreliable, trivial, and unworthy of serious consideration, much less application. (Tate, 1950, 11)

> . . . much that is called research cannot be considered such when gauged by scientific standards. (Fehr, 1950, 11)

> It should be apparent to the reader that few of the studies which have been reported in this review offer evidence which can be accepted without considerable reservation. Many of them are faulty, either in design or in interpretation, or in both. (Johnson, 1944, 482)

Since research efforts vary widely in quality, the question of how much confidence can be placed in the findings of a study is one of considerable importance. Because of this, a comprehensive compilation must contain some indication of the value of each study. This attempt to evaluate is a significant characteristic of the present study.

Review of Related Literature

The literature was searched to find the answers to three questions: (1) Is there a compilation of the research on elementary school mathematics? (2) Is there an evaluative compilation of the research on elementary school mathematics? (3) Is there an instrument for evaluating research?

Previous Compilations of the Research on Elementary School Mathematics. There are many compilations, no one of which is truly comprehensive. The existing ones may be grouped into three classifications: reviews, topical summaries, and bibliographical listings.

Among the reviews are twelve published by the Review of Educational Research. These are descriptive accounts, primarily concerned with reporting significant findings, conclusions, and implications of the research within a specified period of time. The Cyclopedia of Education and the editions of the Encyclopedia of Educational Research contain summaries of the most significant conclusions and implications of research over the years, within the framework of usefulness to teaching, with criteria determined by the reviewer's philosophy. Glennon and Hunnicutt (1952, 1958), Morton (1953), and Spitzer (1962) have discussed the implications of the

research in mathematics for the classroom teacher in pamphlets which enumerate applications but do not directly quote the research. All of the reviews are summarized on Table I in Appendix A.

In the group of topical summaries are those studies which review the research on a particular topic. Several are particularly good examples of carefully done research of this type. Brownell and others (1941) critically analyzed the research for the findings applicable to the teaching of arithmetic in grades 1 and 2. Johnson (1944) compared and noted weaknesses of research on problem solving. The research from 1911 to 1940 on methods of teaching arithmetic was compared by Knipp (1944). A list of these and similar studies is presented on Table II in Appendix A.

Cited as bibliographical are those studies which have as a primary purpose the listing of references. In some cases these are complete for a specified period of time, while in other cases they are selected by criteria not always specified by the reviewer.

Buswell and Judd (1925), recognizing the need for synthesis, compiled a list of 320 titles which included both research and critical discussions. Buswell continued this practice for the next seven years, but from 1933 changed his emphasis, presenting only "selected references" without attempting completeness. Hartung continued this practice from 1943 through 1964.

Monroe and Engelhart (1931) used the lists developed by Buswell and Judd (1925) and Buswell (1926-1930) as a primary source of titles for their summary, which includes only research.

Stretch (1941), Van Engen (1950), Gibb (1954), Hunnicutt and Iverson (1958), and Schaaf (1960) presented selected references, with the basis for selection generally one of pertinence, but not precisely defined.

Weaver (1957, 1958-66) probably presented the most complete lists of what he designated as "normative and experimental research." His sources were journals and other publications, as well as Dissertation Abstracts. He attempted to secure an exhaustive listing, but only for the years since 1950. Some "significant articles," in addition to research reports, were cited. Annotations have been included on several of these lists, but categorization was not a consistent feature.

The compilation of doctoral dissertations by Summers and others (1961, 1963, 1965) would seem to be complete, but is not categorized in all cases. Brown and various co-authors (1953, 1954, 1955, 1958, 1960, 1963, 1965) made little attempt to be exhaustive, but listed current research at the doctoral level and that which was supported by government funding. Burns and Dessart (1965, 1966) summarized investigations for a limited period.

Table III in Appendix A summarizes the most pertinent bibliographical listings.

Previous Evaluative Compilations of the Research on Elementary School Mathematics. Most of the compilations are evaluative only by selection or omission, primarily by the criterion of appropriateness to the specified topic or period of time. In only ten cases are critical comments of some type made. Bernstein (1959) stated evaluative reactions to some of the research on remedial arithmetic. Brownell and others (1941) included critical comments in the course of their discussion on primary arithmetic. Criticism on design and findings was made by Hightower (1954). Writing in 1914, Howell noted that some experiments are open to question. Johnson (1944) pointed out weaknesses in some of the studies on problem solving, as did Weaver (1956) in his critical review of research on compound subtraction. Buswell and Judd (1925), Schaaf (1960), and Weaver (1957,

1958-1966) all noted specific criteria for acceptance or rejection. Monroe and Engelhart (1931) used a criterion to select experimental and research studies primarily from the lists compiled by Buswell and Judd (1925) and Buswell (1926-1930). Four other criteria were utilized to evaluate this research: control of variables, accuracy and validity of the measures used, and justification for the generalization.

Previous Instruments for Evaluating Educational Research. Six instruments for evaluating research, all of which have been tested for reliability, have been found. For three of these (Cook, 1964; Hodges, 1966; Wandt, 1965), no reliability data are available. For one (Shaycroft and Altman, 1955) the reliability is so low (.186) that usefulness of the instrument is questionable. The remaining two have been found to be helpful in evaluating educational research.

Johnson (1957) attempted to "devise and evaluate a technique which would facilitate the acquisition of skill in summarizing and evaluating scientific research articles in education." The technique and the report and evaluation sheet were developed during a graduate course on educational research methods. The interrelationship between student evaluations was .76 (significant beyond .01), using a "random split corrected for attenuation." The relationship between student evaluation and expert evaluation was .78 (significant beyond .01), while the interrelationship between expert evaluations was .79 (significant beyond .01).

Gephart (1964) attempted to "determine the interrater reliability of a research evaluation instrument . . . structured . . . through the identification of action verbs and the objects of these action verbs used in describing the research process." The interrater reliability

for overall evaluation ratings was .76 for rankings and .74 for ratings (significant beyond .001), using Kendall's W.

In addition to these instruments, many suggestions have been made on ways to do better research. Brooks (1923), Brownell (1947), Farquhar and Krumboltz (1959), Fox (1958), Gates (1949), Good (1929, 1963), Kerlinger (1964), McCall (1923), MacDonald (1966), Monroe and Engelhart (1931), Mouly (1963), Perdew (1950), Scates and Hoban (1937), Symonds (1956), Travers (1964), Tyler (1958), Van Dalen (1958), Wolfle (1949), the Encyclopedia of Educational Research, and the Bureaus of Educational Research at the University of Minnesota and the Ohio State University have all suggested criteria for the evaluation of educational research, either in the form of a list or as a specific suggestion.

No evaluative instrument has been applied to the research in elementary school mathematics as far as can be ascertained from the literature.

Description of the Study

A list of all reports of research which relate to the teaching of mathematics in the elementary school (kindergarten through grade eight) and which have been printed in journals published in the United States during the years from 1900 through 1965 has been compiled. Each study was categorized by mathematical topic and type of study. The research which is experimental was also categorized by design paradigm. Specific information on statistical procedure, variables controlled, sampling procedures and size, type of test, grade level, and duration were included whenever applicable. Major conclusions which appear consistent with the data of each study were noted.

An Instrument for Evaluating Experimental Research Reports was developed and tested for reliability. The experimental research was evaluated with this instrument, and each study was assigned to a composite evaluative category.

In addition, a list of dissertations which have been completed has been compiled, in order to increase the comprehensiveness of the compilation.

Pertinent data have been summarized and major conclusions pertaining to mathematical and educational research methodology are reported. Limitations of the study are noted in the dissertation.[1]

---

[1] Suydam, Marilyn N., "An Evaluation of Journal-Published Research Reports on Elementary School Mathematics, 1900-1965," pp. 5-9, Unpublished doctoral dissertation, The Pennsylvania State University, 1967.

II. METHOD

This study involved five stages: (1) compiling, (2) categorizing, (3) developing an instrument, (4) evaluating, and (5) summarizing the data on the reports of research on elementary school mathematics.

Compiling

Reports of research on elementary school mathematics printed from 1900 through 1965 in journals published in the United States have been collected. To identify the reports, several procedures were used to ensure as complete a listing as possible. The journals in which over eighty per cent of the research reports for the post-1930 period appeared were checked on a page-by-page basis. For other post-1930 journals and for all journals in the pre-1930 period, only references cited by others were included. All issues of the Education Index from the first issue in 1929 through 1966 were searched and each reported article which seemed to pertain to research was individually checked. Each report of research was scrutinized for any references made to previous research. In addition, collections of research were examined.

A list of the dissertations on elementary school mathematics which were completed from 1900 through 1965 was compiled to extend the compilation. Dissertation Abstracts and previous investigations provided the major sources for this list.

Categorizing

Each report of journal-published research was categorized by mathematical topic (see Appendix D) and type of study (see Appendix E).

Experimental research was categorized by design paradigm, statistical procedure, sampling procedure and size, type of test, grade level, and duration. When such information was available in a report for another type of research, it was included. The major conclusions or findings which seemed supported by the data have been noted for all types of studies, and the independent (I) and dependent (D) variables have been noted for experimental research, and for those types of action research where it was possible to do so.

Developing an Instrument

While there is much in the literature on the need to evaluate research, there is comparatively little specific help. Only two instruments were readily available which were considered for use in evaluating the research in arithmetic. However, neither of these proved entirely suitable for the purpose. More information seemed to be needed to support the items on the list developed by Johnson (1957).[2] Gephart (1964) supplied additional information, but then sacrificed the careful time-consuming rating of each sub-item to a purely subjective final rating.[3]

There is a need for a comparatively simple instrument which provides information concerning major factors and problems in research. Thus, the more carefully controlled research can be separated from that which was less well done. It is difficult, however, to attempt to distinguish weaknesses in the research process from those of the reporting process. Therefore, it is more precise to consider the

---

[2,3] See pages 6-7 for additional comments on these instruments.

result as an evaluation of a report; however, its correlation with that of the research on which it is based should be high.

The first stage in the formulation of the Instrument for Evaluating Experimental Research Reports was to compile the lists of suggestions proposed by writers in the field of educational research.[4] The following topics were consistently listed:

(1) Importance or significance of the problem

(2) Definition of the problem

(3) Design of the study

(4) Control of variables

(5) Sampling procedures

(6) Use of instruments

(7) Analysis of data

(8) Interpretation of results

(9) Reporting of the research.

Each of the points was stated in question form, to make it possible to consider an evaluation. The nine questions were checked for completeness of content, and were subjected to trial use in evaluating several reports. It was evident that the instrument could be used more effectively if some direction could be given in answering the nine questions. Using "key points" with adjectives to give a range for each made it reasonably certain that each rater would be focusing upon the same aspect.

---

[4] For a complete list of these, see page 7.

The instrument was tested for interrater reliability; the reports of this testing are included in Chapter III. The instrument itself and directions for its use are presented in Appendix F.

Evaluating

The compiled research published from 1900 through 1965 which was categorized as "experimental" was evaluated with the Instrument for Evaluating Experimental Research Reports. The restriction to experimental research was necessary because of the design of the instrument.

A quantitative score was derived from evaluation with this instrument. The sum of the numerical scores assigned to each question may be considered as a basis for some degree of comparison.

As a final index to the research, each of the research studies was assigned to a composite evaluative category. This index is included to aid the reader in locating those studies which may best meet his purposes. Symbols were chosen to represent:

EPD - Purpose, type of study, design, and statistical procedures seem sound and pertinent to curriculum today under the stated definition of experimental research.

ED - Type of study, design, and statistical procedures seem sound and pertinent to curriculum today under the stated definition of experimental research, but the purpose does not seem pertinent.

EP - Purpose seems pertinent to curriculum today, but type of study, design, and/or statistical procedures do not seem sound and/or accurate today under the stated definition of experimental research.

NE - Study is not considered experimental research under the stated definition.

Summarizing

The results of this study are summarized through the presentation of pertinent data. The total number of uses within each of the categories is depicted on tables. Major conclusions are cited, and the major repetitive errors in research methodology are indicated. Implications for further research are noted.

III. RESULTS

Investigations of the Reliability of the Instrument

The Instrument for Evaluating Experimental Research Reports which was developed as one aspect of this study was tested in two separate investigations for the degree of reliability or interrater agreement which could be expected in its use.

The first study was on a smaller scale than the second, and was conducted prior to any evaluation of the research reports. The population of studies was limited by the extent to which the compilation of reports had proceeded. Its purpose was to ascertain the level of agreement among the writer and two other raters with a comparable background.[5] Provision was made for testing the effect of bias in reading research reports. Since the period of training for this study was limited, the measure of reliability secured may be considered to depict a base level, rather than one inflated by the results of training which almost always would lead to increased agreement.

In the second study, there was no training provided beyond the directions stated on the instrument, for the same reason as in the first study. Moreover, this provides a measure of the usefulness of the instrument to diverse readers of a type who might plausibly use the instrument in a realistic situation without extensive training in its use.

---

[5] These two raters, Cecil R. Trueblood and Lynn A. Watson, aided the writer in evaluating the research for the years 1955 through 1965.

First Study. The procedures for the first study of interrater agreement were:

1. The population of reports of experimental research which have been published in The Arithmetic Teacher from 1954 through 1965 (volumes 1 through 12) was identified.
2. A sample of ten of these reports was randomly selected for reproduction.
3. The name of the author and the year of publication were deleted on five of the reproduced reports, selected at random.
4. Three doctoral candidates in elementary education were identified as raters.
5. The raters independently evaluated the ten selected reports of experimental research with the proposed instrument.

The interrater agreement on overall ratings was determined, using an analysis of variance procedure. The results for the first study of reliability are presented on Table IV in Appendix B.

To determine the proper terms to use in the reliability formula, expected mean squares were determined. Pooling was necessary to secure error terms. When it was assumed that A was a fixed factor, and B and C were random, the only significant effect was that for between Articles (B within A). The interpretation of generalizability is thus extended to include all judges of the same type, though with recognition of the fact that power is lacking due to the small number of judges involved in the study. The F ratios are presented on Table V in Appendix B.

Since the masking treatment (A) and Between Judges (C) effects were non-significant, the proper terms to use in the AOV formula to obtain the coefficient of rater agreement are:

$$r = 1 - \frac{MS_{\text{pooled error term}}}{MS_b}$$

With data from the present study inserted, the result is a coefficient of .91 for interrater agreement.

$$r = 1 - \frac{19.5}{218.0} = .91$$

This coefficient estimates the correlation between the combined ratings of the three judges used in the study and the combined rating of another hypothetical random sample of judges taken from the same population and rating the same ten articles.

The measures of reliability for two previously cited instruments were obtained by different means. Johnson (1957), who found reliability coefficients of .75 for student evaluations and .79 for ratings by experts, used a "random split corrected for attenuation." Gephart (1964) secured an interrater reliability of .764 for rankings and .749 for ratings using Kendall's W. The use of the AOV reliability formula with the data from the present study is somewhat comparable to the statistical treatments which were used in previous studies. The coefficient of .91 compares favorably with these other estimates of observer agreement.

However, as Ebel (1951, 408) states, this formula is sometimes inappropriate when interrater agreements are in question:

> If decisions are based upon average ratings, it of course follows that the reliability with which one should be concerned is the reliability of those averages. However, if the raters ordinarily work individually, and if multiple scores for the same theme or student are only available in experimental situations, then the reliability of individual ratings is the appropriate measure.

He suggests the use of an intraclass formula such as that presented by Snedecor for the reliability of individual ratings:

$$r = \frac{MS_b - MS_r}{MS_b + (k-1)\,MS_r}$$

With data from this study inserted this results in the following:

$$r = \frac{218.0 - 19.5}{218.0 + 2\,(19.5)} = .77$$

Thus the coefficienr of reliability which provides a measure of the consistency probable with a single rater using the Instrument for Evaluating Experimental Research Reports was found to be .77 in this study. This is similar to the coefficients found for previous instruments with less rigorous formulas. As a cross-check on the accuracy of this result, the interrater correlations were found: $r_{1,\,2} = .79$, $r_{2,\,3} = .69$, $r_{1,\,3} = .86$. The mean of the interrater correlations is .78, which confirms the accuracy of the measure of intraclass reliability. It would seem that this coefficient is highly satisfactory.

Second Study. The procedures for the second study paralleled those for the first study except that a more diverse population was considered:

1. A population of reports of experimental research which have been published in journals in the United States from 1930 through 1965 was identified.

2. A stratified sample of ten of these reports was selected. Stratification was on the basis of a) journal source, b) status of author, and c) year of publication.

3. These ten reports were reproduced. On five, selected at random, journal source, status of author, and year of publication were deleted.

4. Twelve raters who were representative of groups most likely to be involved in evaluating educational research were identified:

   a. Three doctoral candidates in elementary education

   b. Three doctoral candidates in educational psychology

   c. Three professors in elementary education

   d. Three professors in educational psychology.

5. Each of the raters independently evaluated each of the ten reports with the proposed instrument.

The results of the analysis of variance for the second study are presented on Table VI in Appendix B.

Expected mean squares were determined for the condition where articles (BwA) and judges (EwCD) were both random, and all other factors fixed. The F ratios for the second study are presented on Table VII in Appendix B. Other effects were nonsignificant; they were pooled to form the error term.

It will be noted from the table that the articles effect (BwA) was significant, as in the first study. Therefore generalizability may be considered to extend to all judges in these subsets, though this study provides more power.

Using the AOV reliability formula, the coefficient of .94 for interrater agreement was found.

$$r = 1 - \frac{11.08}{183.97} = .94$$

When Snedecor's formula is used, the resulting coefficient is .57.

$$r = \frac{183.97 - 11.08}{183.97 + (11)\ (11.08)} = .57$$

When correlations between each pairing of the twelve raters were computed, the mean was found to be .57. This serves as a check on the accuracy of the intraclass reliability.

The degree of interrater agreement again compares favorably with those found for previous instruments. The reliability of individual ratings which is derived from Snedecor's formula is lower, and may present a more realistic picture of the variability which may be expected from the use of a rating instrument of this kind with a single rater.

As a comparative analysis for interest and information, each of the four subsets of three raters was considered separately. The coefficients for interrater agreement and for individual reliability which were found for each subset of raters are presented on Table VIII in Appendix B.

Implications. The Instrument for Evaluating Experimental Research Reports was found to have coefficients of interrater agreement which ranged from .77 to .94, using the analysis of variance formula. The particular set of judges being considered caused the range. The reasons for the range with varying groups are a matter of conjecture,

and beyond the scope of the present study. The set of judges was apparently quite homogeneous, since there is apparently one general factor which is being tested, with all other factors accounting for only a small portion of the variance.

The measures of reliability reported were based on studies in which the raters received no training. In previous investigations with similar scales, it has been found that training will increase the degree to which raters agree.

The coefficients for one judge using the instrument ranged from .53 to .78. Any one individual's perception apparently lowers the level of reliability which can be predicted. Anyone who uses the instrument should ascertain the degree of interrater agreement and/or the coefficient for one rater which applies to that particular situation.

Summarization and Analysis of Data

A total of 799 analyses are presented in the dissertation.[6]

Journals. These 799 research reports were found in fifty journals presented in Table IX in Appendix C. Three journals published over half (54%) of the reports. Ten journals published 84% of the reports; thirteen journals, 89%. The remaining reports (11%) were published in 37 journals.

Years. A count of the distribution by years revealed that 2 reports were found for the decade 1900-1910; 36 for 1911-1920; 89 for 1921-1930; 167 for 1931-1940; 118 for 1941-1950; 165 for 1951-1960; and 222 for 1961-1965. The figure for the last five-year period is obviously greater than for any prior ten-year period, underlining the emphasis being placed on research today.

---

[6] Suydam, Marilyn N., op. cit., pages 50-438.

Mathematical topic and type of study. Table X in Appendix C presents the frequency by mathematical topic and the frequency by type of study. The number of reports of experimental research was 246, a figure almost equalled by the 230 reports of surveys which were found. Totals for other types of studies were: descriptive, 107; case study, 18; action, 63; correlational, 56; and ex post facto, 79.

The distribution of reports gives some indication of the concern for various topics, as well as an indication of the fact that some topics lend themselves more readily to one type of research. For instance, readiness (b-1) is most readily ascertained through surveys, while case studies were most frequently used to depict individualization techniques, particularly for remediation (e-2).

Cross-referencing. Cross-referencing adds more depth, for in many cases the topic which was cited first was selected arbitrarily. The totals within each mathematical category shift somewhat as all references are counted. The topics under which the largest number of all types of research were categorized are:

(1) a-5b: problem solving (84)

(2) f-2: achievement evaluation (76)

(3) a-3: planning and organizing for teaching (62)

(4) d-1: textbooks (56)

(5) e-1: diagnosis (55)

(6) e-2: remediation (52)

(7) b-5: content to be included in grade (46)

(8) f-1: testing (44)

(9) a-5a: drill and practice (43)

(10) c-8: measurement (43)

(11) c-3d: division of whole numbers (40)

Design paradigm. A frequency distribution was made for the design paradigms which were categorized. Those more frequently noted were:

(1) 3.4: pretest-posttest, insufficient information re n (50)

(2) 1.2: one group pretest-posttest (25)

(3) 3.21: non-equivalent control group, pretest-posttest (18)

(4) 3.22: non-equivalent control group, posttest only (18)

(5) 3.19: posttest only, own control, insufficient information re n (17)

(6) 2.2: pretest-posttest, control group, matched, n = students (14)

(7) 3.1: pretest-posttest, control group, matched, n = students when the sampling unit seems to have been classes (13)

(8) 3.8: posttest only, insufficient information re n (13)

Analysis of these types reveals a problem which is shown in several ways: sampling and/or the way in which a researcher reported the sampling for his experiment was a point of great variability and ambiguity. Of the 246 experimental studies, 39 involved no control group, while another 150 involved possible sampling errors.

Statistical procedure. Descriptive statistics are noted in almost 2/3 of the reports. The other techniques most noted were:

(1) 3.4: t-test (123)

(2) 6.4: correlation (89)

(3) 3.3: F-test (68)

(4) 3.2: analysis of variance (58)

(5) 3.15: z-test, critical ratio (43)

(6) 3.5: analysis of covariance (37)

(7) 2.6: Chi square test for independence (30)

(8) 3.17: Probable error (24)

Evaluative category. The final evaluative category was included as a referrent for determining ultimate value of the studies in the opinion of the reviewer. The majority of the studies (553) were labeled non-experimental. Of the 246 experimental studies, 112 were labeled "EPD"[7]; 9 were labeled "ED"; and 125 were labeled "EP". Thus, only 14% of all studies or 46% of the experimental studies were considered sound and pertinent today as experimental research.

Qualitative value. Analysis of the qualitative values which resulted from application of the Instrument For Evaluating Experimental Research Reports shows a range from 13 to 44 of a possible 9 to 45. Table XI in Appendix C shows the distribution by three periods of time: 1900-1929, 1930-1950, and 1951-1965.

Two questions, those involving control of variables and sampling, were rated especially low. The percentages for those which were assigned ratings of satisfactory or better on each question are:

(1) How practically or theoretically significant is the problem? 73.5%

(2) How clearly defined is the problem? 72.3%

(3) How well does the design answer the research question? 50.7%

---

[7] See pages 12 and 13 for definitions of these symbols.

(4) How adequately does the design control variables? 29.7%

(5) How properly is the sample selected for the design and purpose of the research? 27.7%

(6) How valid and reliable are the measuring instruments or observational techniques? 53.3%

(7) How valid are the techniques of analysis of data? 44.4%

(8) How appropriate are the interpretations and generalizations to the data? 59.8%

(9) How adequately is the research reported? 65.0%

Dissertations. A total of 470 dissertations on elementary school mathematics were found for the 65-year period.

Analysis of Content. Only eighty reports of the 246 in the experimental category were considered satisfactory or better on total scores. This would seem to indicate a need to improve the reporting of research and possibly research procedures as well.

When these eighty studies are considered, no possible summary can be made, either because there was only one study in a category or because the studies were aimed at diverse phases. In other cases, inconsistency is evidenced. Some specific help is provided for the classroom teacher--and this is the ultimate purpose of any research--but there is no clear and well-defined pattern evidenced from resea

IV. SUMMARY AND IMPLICATIONS

Final Summary

1. A list of all reports of research which relate to the teaching of mathematics in the elementary school and which have been printed in journals published in the United States during the years from 1900 through 1965 has been compiled. A total of 799 research reports were found in 50 journals.

2. Each study is categorized by mathematical topic and type of study. Of the total, 207 were placed primarily in the categories for educational objectives and instructional procedures; 63 in topical placement; 154 in basic concepts and methods of teaching them; 78 in materials; 131 in individual differences; 99 in evaluating progress; and 67 were categorized as studies related to learning theory. The frequency by types of studies was: descriptive, 107; survey, 230; case study, 18; action, 63; correlational, 56; ex post facto, 79; and experimental, 246.

3. The research which is experimental is also categorized by design paradigm. Of the 246 experimental studies, 39 involved no control group; 150 involved possible sampling errors; while only 57 seemed to be valid examples of more carefully designed experiments.

4. Specific information on statistical procedure, variables controlled, sampling procedures and size, type of test, grade level, and duration are included whenever applicable in the analysis of each report.

5. Major conclusions which appear consistent with the data in each study are also noted with the analysis of each report.

6. An Instrument for Evaluating Experimental Research Reports was developed and tested for reliability. In one study with three judges, the interrater agreement was found to be .19, while the intraclass reliability was .77. In a second study with twelve judges, the interrater reliability was found to be .94, with an intraclass reliability coefficient of .58.

7. The experimental research is evaluated with this instrument. None of the reports was rated excellent in overall rating. 34 of the reports were rated very good; 60, good; 84, fair; and 68, poor.

8. Each study was assigned to a composite evaluative category. 553 were non-experimental; 112, "EPD" (purpose, type of study, design, and statistical procedures seem sound and pertinent to curriculum today under the stated definition of experimental research); 9, "ED" (type of study, design, and statistical procedures seem sound and pertinent to curriculum today under the stated definition of experimental research, but the purpose does not seem pertinent); and 125 "EP" (purpose seems pertinent to curriculum today, but type of study, design, and/or statistical procedures do not seem sound and/or accurate today under the stated definition of experimental research).

9. A list of 470 dissertations which have been completed was compiled and included in the appendix to increase the comprehensiveness of the compilation.

10. Pertinent data were summarized and major conclusions pertaining to mathematical and educational research methodology are listed. No clearly defined applicability to a theory of instruction is evident.

Suggestions and Implications

The first time a project of the type involved in the present study is attempted, there is a process of evolvement. The basic model or structure is revised time and again. Therefore, the following suggestions are made:

1. There is a need to replicate this study, using the present structure as the basis.

2. More precise definitions of such categories as design paradigm can be more readily developed now that a firmer concept exists of what is actually found in the research on elementary school mathematics. Through such precisely defined categories, the factor of perceptual differences may be more readily controlled.

3. The statistical procedure and other categories could be checked for accuracy.

4. More extensive cross-referencing could be done.

5. The research could be re-evaluated to secure a confirmation of the validity of the present evaluation.

Continued extension to add reports of research on elementary school mathematics is necessary. This would include:

1. Reports published in American journals for the period 1900

through 1965 which were not discovered by the present reviewer need to be included in the compilation.

2. Other sources of research reports need to be searched and compilations developed.

3. Confirmation of the accuracy of the list of dissertations is needed.

4. The compilation and the evaluation of the reports need to be extended beyond 1965.

Synthesis of the data in a form which is meaningful to teachers should be done. In particular, an analysis of what the research should and can mean to classroom teachers is of vital importance.

It was noted that two major deficiencies are evident in research reports: (a) the lack of sufficient information on sampling and (b) the lack of firm control of variables. These may be merely problems of reporting. They may also be actual problems of the research process. Thus it seems that:

1. The improvement of research possibly depends on increasing the researcher's awareness of the need to consider these two points especially carefully.

2. The evaluation with the other seven points on the instrument would seem to show that more careful planning and reporting of research projects are needed.

3. The Instrument for Evaluating Experimental Research Reports may serve as a guide to planning as well as its use in evaluating the finished product.

4. There is a need to develop similar instruments to evaluate types of research other than experimental. It is the opinion of the writer that sampling was a problem in most types of research.

5. Researchers need to consider the possibility of planning experimental research rather than, as has happened in the past, resorting to ex post facto studies.

6. Careful and precise planning of research is vital. Equally careful and precise reporting would be helpful.

More research needs to be done on many topics. The topics of the 799 studies seemed to be almost randomly distributed among categories. Researchers need to consider the points on which research is most needed. The possibility of using research as a means of developing a theory of instruction needs to be carefully and thoughtfully pursued.

LIST OF REFERENCES

A. References Related to Research on Elementary School Mathematics

Beatty, Leslie S. Re-Orienting to the Teaching of Arithmetic. Childhood Education. 26: 272-278; February, 1950.

Bernstein, Allen L. Library Research -- A Study in Remedial Arithmetic. School Science and Mathematics. 59: 185-195; March, 1959.

Brown, J. C. A Summary of Some Significant Conclusions Reached by Investigators Relative to Arithmetic. Elementary School Journal. 25: 346-357; January, 1925.

Brown, Kenneth E. Mathematics Education Research Studies (1952-1954). Circular No. 377-I, II, III. Washington, D.C.: Government Printing Office, 1953, 1954, 1955.

Brown, Kenneth E. Research in Mathematics Education. Mathematics Teacher. 46: 594-597; December, 1953.

Brown, Kenneth E. Research in Mathematics Education: 1953. Mathematics Teacher. 48: 21-25; January, 1955.

Brown, Kenneth E. Analysis of Research in the Teaching of Mathematics, 1955 and 1956. U. S. Dept. of Health, Education and Welfare, Office of Education, Bulletin No. 4. Washington, D.C.: Government Printing Office, 1958.

Brown, Kenneth E., and Abell, Theodore L. Analysis of Research in the Teaching of Mathematics, 1965. U. S. Dept. of Health, Education and Welfare, Office of Education, Bulletin No. 28. Washington, D.C.: Government Printing Office, 1965.

Brown, Kenneth E., and Abell, Theodore L. Research in the Teaching of Elementary School Mathematics. Arithmetic Teacher. 12: 547-549; November, 1965.

Brown, Kenneth E., and Kinsella, J. J. Analysis of Research in the Teaching of Mathematics, 1957 and 1958. U. S. Dept. of Health, Education and Welfare, Office of Education, Bulletin No. 8. Washington, D.C.: Government Printing Office, 1960.

Brown, Kenneth E., and others. Analysis of Research in the Teaching of Mathematics, 1959 and 1960. U. S. Dept. of Health, Education and Welfare, Office of Education, Bulletin No. 12. Washington, D.C.: Government Printing Office, 1963.

Brownell, William A. Readiness and the Arithmetic Curriculum. Elementary School Journal. 38: 344-354; January, 1938.

Brownell, William A. Teaching Mathematics in Grades I through VI. Review of Educational Research. 15: 276-288; October, 1945.

Brownell, William A., and Grossnickle, Foster E. Teaching Mathematics in Grades I through VI. Review of Educational Research. 12: 386-404; October, 1942.

Brownell, William A., and others. Arithmetic in Grades I and II. Duke University Research Studies in Education, No. 6. Durham, N.C.: Duke University Press, 1941.

Brueckner, Leo J. Arithmetic. Review of Educational Research. 4: 140-143, 215-217; April, 1934.

Brueckner, Leo J., and others. Arithmetic. Review of Educational Research. 7: 453-463, 545-547; December, 1937.

Burch, Robert L., and Moser, Harold E. The Teaching of Mathematics in Grades I through VIII. Review of Educational Research. 21: 290-304; October, 1951.

Burns, Paul C. Arithmetic Research That Has Made a Difference. Elementary School Journal. 65: 386-392; April, 1965.

Burns, Paul C., and Dessart, Donald J. A Summary of Investigations Relating to Mathematics in Elementary Education: 1964. School Science and Mathematics. 65: 779-790; December, 1965.

Burns, Paul C., and Dessart, Donald J. A Summary of Investigations Relating to Mathematics in Elementary Education: 1965. School Science and Mathematics. 66: 838-851; December, 1966.

Buswell, Guy T. Summary of Arithmetic Investigations. Elementary School Journal. 26: 692-703, 745-758; May, June, 1926. 27: 685-694, 731-744; May, June, 1927. 28: 702-709, 730-742; May, June, 1928. 29: 691-698, 737-747; May, June, 1929. 30: 766-775; June, 1930. 31: 756-766; June, 1931. 32: 766-773; June, 1932.

Buswell, Guy T. A Critical Survey of Previous Research in Arithmetic. In Report of the Society's Committee on Arithmetic. Twenty-ninth Yearbook of the National Society for the Study of Education. Bloomington, Ill.: Public-School Publishing Co., 1930. pp. 445-470.

Buswell, Guy T. Selected References on Elementary School Instruction -- Arithmetic. Elementary School Journal. 34: 209-213;

November, 1933. 35: 212-214; November, 1934. 36: 211-213; November, 1935. 37: 209-212; November, 1936. 38: 210-214; November, 1937. 39: 211-213; November, 1938. 40: 212-214; November, 1939. 41: 210-212; November, 1940. 42: 212-215; November, 1941. 43: 174-175; November, 1942. 44: 167-168; November, 1943. 45: 164-165; November, 1944. 46: 160-162; November, 1945. 47: 161-163; November, 1946. 48: 155-157; November, 1947.

Buswell, Guy T. Arithmetic. Encyclopedia of Educational Research. (Edited by C. W. Harris.) New York: Macmillan Co., 1960. pp. 63-77.

Buswell, Guy T., and Judd, Charles H. Summary of Educational Investigations Relating to Arithmetic. Supplementary Educational Monographs, No. 27. Chicago: University of Chicago Press, 1925.

Dawson, Dan T., and Ruddell, Arden K. The Case for the Meaning Theory in Teaching Arithmetic. Elementary School Journal. 55: 393-399; March, 1955.

Fehr, Howard F. Present Research in the Teaching of Arithmetic. Teachers College Record. 52: 11-23; October, 1950.

Fisher, S. Carolyn. Arithmetic and Reasoning in Children. Pedagogical Seminary. 19: 48-77; March, 1912.

Gane, James T. Research Should Guide Us. Arithmetic Teacher. 9: 441-445; December, 1962.

Gibb, E. Glenadine. A Review of a Decade of Experimental Studies Which Compared Methods of Teaching Arithmetic. Journal of Educational Research. 46: 603-608; April, 1953.

Gibb, E. Glenadine. A Selected Bibliography of Research in the Teaching of Arithmetic. Arithmetic Teacher. 1: 20-22; April, 1954.

Gibb, E. Glenadine, and Van Engen, Henry. Mathematics in the Elementary Grades. Review of Educational Research. 27: 329-342; October, 1957.

Glennon, Vincent J., and Hunnicutt, C. W. What Does Research Say About Arithmetic? Washington, D.C.: Association for Supervision and Curriculum Development, National Education Association, 1952, 1958.

Greene, Harry A. A Critique of Remedial and Drill Materials in Arithmetic. Journal of Educational Research. 21: 262-276;

April, 1930.

Hartung, Maurice L. Selected References on Elementary-School Instruction -- Arithmetic. Elementary School Journal. 49: 168-170; November, 1948. 50: 165-167; November, 1949. 51: 157-159; November, 1950. 52: 162-164; November, 1951. 53: 163-165; November, 1952. 53: 163-165; November, 1953. 55: 162-164; November, 1954. 56: 124-125; November, 1955. 57: 161-162; December, 1956. 58: 170-172; December, 1957. 59: 170-172; December, 1958. 60: 166-168; December, 1959. 61: 167-170; December, 1960. 62: 165-167; December, 1961. 63: 170-172; December, 1962. 64: 169-171; December, 1963. 65: 163-166; December, 1964.

Hartung, Maurice L. Estimating the Quotient in Division (A Critical Analysis of Research). Arithmetic Teacher. 4: 100-111; April, 1957.

Hightower, H. W. Effect on Instructional Procedures on Achievement in Fundamental Operations in Arithmetic. Educational Administration and Supervision. 40: 336-348; October, 1954.

Hunnicutt, C. W., and Iverson, William J. (Editors). Research in the Three R's. New York: Harper & Brothers, 1958.

Johnson, Harry C. Problem-Solving in Arithmetic: A Review of the Literature. Elementary School Journal. 44: 396-403, 476-482; March, April, 1944.

Knipp, Minnie B. An Investigation of Experimental Studies Which Compare Methods of Teaching Arithmetic. Journal of Experimental Education. 13: 23-30; September, 1944.

Monroe, Walter S. and Engelhart, Max D. A Critical Summary of Research Relating to the Teaching of Arithmetic. Bureau of Educational Research Bulletin No. 58. Urbana: University of Illinois, 1931.

Morton, Robert L. Teaching Arithmetic. What Research Says to the Teacher, No. 2. Washington, D.C.: National Education Association, 1953.

Moser, Harold E., and others. Aims and Purposes in the Teaching of Mathematics. Review of Educational Research. 18: 315-322; October, 1948.

Payne, Holland. What About Modern Programs in Mathematics? Mathematics Teacher. 58: 422-424; May, 1965.

Payne, Joseph N., and Goodman, Frederick L. Mathematics Education, Key-Word-in-Context Index and Bibliography, Volumes I and II. Ann Arbor, Mich.: Campus Publishers, 1965.

Pikal, Frances. Review of Research Related to the Teaching of Arithmetic in the Upper Elementary Grades. School Science and Mathematics. 57: 41-47; January, 1957.

Riess, Anita. Number Readiness in Research (A Survey of the Literature). Chicago: Scott, Foresman & Co., 1947.

Ruch, G. M., and Mead, Cyrus D. A Review of Experiments on Subtraction. In Report of the Society's Committee on Arithmetic. Twenty-ninth Yearbook of the National Society for the Study of Education. Bloomington, Ill.: Public-School Publishing Co., 1930. pp. 671-678.

Schaaf, William L. Selected Annotated Bibliography. In Instruction in Arithmetic. Twenty-fifth Yearbook of the National Council of Teachers of Mathematics. Washington, C.D.: The Council, 1960. pp. 320-354.

Schreiber, Margaret. A Summary of Arithmetic Research. Educational Administration and Supervision. 34: 368-372; October, 1948.

Sherer, L. Some Implications from Research in Arithmetic. Childhood Education. 29: 320-324; March, 1953.

Smith, David Eugene, and Goldziher, Charles. Bibliography of the Teaching of Mathematics: 1900-1912. U. S. Bureau of Education, Bulletin No. 29. Washington, D.C.: Government Printing Office, 1912.

Smith, David Eugene. Arithmetic; Mathematics. Cyclopedia of Education. (Edited by Paul Monroe.) New York: Macmillan Co., 1918. Volume I: 203-207. Volume II: 159-160.

Spitzer, Herbert F. What Research Says About Teaching Arithmetic. Washington, D.C.: National Education Association, 1962.

Spitzer, Herbert F., and Burch, Robert L. Methods and Materials in the Teaching of Mathematics. Review of Educational Research. 18: 337-349; October, 1948.

Spitzer, Herbert F., and Burns, Paul C. Mathematics in the Elementary School. Review of Educational Research. 31: 248-249; June, 1961.

Stone, M. V. Partial Survey of Studies Published in 1930 in the Field of Arithmetic Methods. Education. 52: 333-336; February, 1932.

Stretch, Lorena B. One Hundred Selected Research Studies. In Arithmetic in General Education. Sixteenth Yearbook of the

National Council of Teachers of Mathematics. Washington, D.C.: The Council, 1941. pp. 318-327.

Summers, Edward G. A Bibliography of Doctoral Dissertations Completed in Elementary and Secondary Mathematics from 1918-1952. School Science and Mathematics. 61: 323-335; May, 1961.

Summers, Edward G. Elementary and Secondary Science and Mathematics Dissertations Reported in 1962. School Science and Mathematics. 63: 733-738; December, 1963.

Summers, Edward G., and Hubrig, Billie. Doctoral Dissertation Research in Mathematics Reported for 1963. School Science and Mathematics. 65: 505-528; June, 1965.

Summers, Edward G., and Stochl, James E. A Bibliography of Doctoral Dissertations Completed in Elementary and Secondary Mathematics from 1950-1960. School Science and Mathematics. 61: 431-439; June, 1961.

Van Engen, Henry. A Summary of Research and Investigations and Their Implications for the Organization and Learning of Arithmetic. Mathematics Teacher. 41: 260-265; October, 1948.

Van Engen, Henry. A Selected List of References on Elementary School Arithmetic. Mathematics Teacher. 43: 168-171; April, 1950.

Weaver, J. Fred. Six Years of Research on Arithmetic Instruction: 1951-1956. Arithmetic Teacher. 4: 89-99; April, 1957.

Weaver, J. Fred. Research on Arithmetic Instruction (1957-1959). Arithmetic Teacher. 5: 109-118; April, 1958. 6: 121-132; April, 1959. 7: 253-265; May, 1960.

Weaver, J. Fred. A Bibliography of Selected Summaries and Critical Discussions of Research in Elementary School Mathematics. Arithmetic Teacher. 7: 364-366; November, 1960.

Weaver, J. Fred. Analysis of Research in the Teaching of Mathematics: 1957 and 1958. Arithmetic Teacher. 8: 81-82; February, 1961.

Weaver, J. Fred. Research on Elementary School Mathematics (1960-1965). Arithmetic Teacher. 8: 255-260, 301-306; May, October, 1961. 9: 287-290; May, 1962. 10: 297-300; May, 1963. 11: 273-275; April, 1964. 12: 382-387, 577-578; May, November, 1965.

Weaver, J. Fred. Research on Mathematics Education, Grades K-8, for 1965. Arithmetic Teacher. 13: 414-427; May, 1966.

Weaver, J. Fred. Differentiated Instruction and School-Class Organization for Mathematical Learning Within the Elementary Grades. Arithmetic Teacher. 13: 495-506; October, 1966.

Weaver, J. Fred, and Gibb, E. Glenadine. Mathematics in the Elementary School. Review of Educational Research. 34: 273-285; June, 1964.

Wheat, Harry G. Research in Teaching Methods of Self-Instruction. Educational Research Bulletin. 20: 87-93, 112; April 16, 1941.

Wheat, Harry G. Studies in Arithmetic. Morgantown, W. Va.: West Virginia University, 1945.

Willey, Roy DeVerl. The Use of Arithmetic in the Out-of-School Life of Children. Mathematics Teacher. 35: 23-28; January, 1942.

Wilson, Guy M. Arithmetic. In Encyclopedia of Educational Research. (Edited by W. S. Monroe.) New York: Macmillan Co., 1941. pp. 42-58.

Wilson, Guy M. The Social Utility Theory as Applied to Arithmetic: Its Research Basis and Some of Its Implications. Journal of Educational Research. 41: 321-337; January, 1948.

Wilson, Guy M. Arithmetic. In Encyclopedia of Educational Research. (Edited by W. S. Monroe.) New York: Macmillan Co., 1950. pp. 44-58.

Woody, Clifford. Arithmetic, Fine Arts, Physical and Health Education, and Industrial Arts. Review of Educational Research. 1: 261-267; 307-310; October, 1931.

Woody, Clifford. Arithmetic. Review of Educational Research. 5: 14-30, 93-97; February, 1935.

Wrightstone, J. Wayne. Influence of Research on Instruction in Arithmetic. Mathematics Teacher. 45: 187-192; March, 1952.

B. References Related to Methodology of Educational Research

Barr, A. S. Research Methods. In Encyclopedia of Educational Research. (Edited by C. W. Harris.) New York: Macmillan Co., pp. 1160-1165.

Brooks, F. D. Criteria of Educational Research. School and Society. 18: 724-729; December 22, 1923.

Brownell, William A. The Techniques of Research Employed in Arithmetic. In Report of the Society's Committee on Arithmetic. Twenty-ninth Yearbook of the National Society for the Study of Education. Bloomington, Ill.: Public-School Publishing Co., 1930. pp. 415-443.

Brownell, William A. Frontiers in Educational Research in Arithmetic. Journal of Educational Research. 40: 373-380; January, 1947.

Brownell, William A. The Evaluation of Learning Under Dissimilar Systems of Instruction. Arithmetic Teacher. 13: 367-374; April, 1966.

Buswell, Guy T. The Outlook for Research in Arithmetic. Elementary School Journal. 47: 243-253; January, 1947.

Campbell, Donald T., and Stanley, Julian C. Experimental and Quasi-Experimental Designs for Research on Teaching. In Handbook of Research on Teaching. (Edited by N. L. Gage.) Chicago: Rand McNally & Co., 1963. pp. 171-246.

Clark, David L; Guba, Egon G.; and Smith, Gerald R. Functions and Definitions of Functions of a Research Proposal or Research Report. Ohio State University. Undated. (ditto)

Cook, Desmond L. An Analysis of Research Methodology Employed in Studies of Programmed Instruction. Ohio State University, 1964. (multilith)

Cook, Desmond L. Checklist for Evaluating Design Aspects of Experimental Research. Ohio State University, 1964. (multilith)

Dvorak, Earl A. General Guide to a Study of Research Reports. Peabody Journal of Education. 34: 141-144; November, 1956.

Ebel, Robert L. Estimation of the Reliability of Ratings. Psychometrika. 15: 407-424; December, 1951.

Farquhar, William W., and Krumboltz, John D. A Check List for Evaluating Experimental Research in Psychology and Education. Journal of Educational Research. 52: 353-354; May, 1959.

Fehr, Howard F. Present Research in the Teaching of Arithmetic. Teachers College Record. 52: 11-23; October, 1950.

Fox, James H. Criteria of Good Research. Phi Delta Kappan. 39: 284-286; March, 1958.

Freeman, Frank N. Research in Perspective. California Journal of Educational Research. 4: 195-200; November, 1953.

Gates, Arthur I. Some Characteristics of Research in the Teaching of Reading. Supplementary Educational Monographs, No. 69. Chicago: University of Chicago Press, 1949.

Gephart, William J. Development of an Instrument for Evaluating Educational Research Reports. Cooperative Research Project No. S-014, U. S. Office of Education, 1964.

Good, Carter V. How To Do Research in Education. Baltimore: Warwick & York, Inc., 1929.

Good, Carter V. Methods of Research and Problem Solving in Education. Journal of Educational Research. 34: 81-89; October, 1940.

Good, Carter V. Introduction to Educational Research. New York: Appleton-Century-Crofts, 1963.

Hodges, Caroline S. Detailed Criteria Rating Form (Revised Version No. 2). New York: Bureau of Applied Social Research, Columbia University, 1966. (multilith)

Johnson, Granville B., Jr. A Method for Evaluating Research Articles in Education. Journal of Educational Research. 51: 149-151; October, 1957.

Kerlinger, Fred N. Foundations of Behavioral Research. New York: Holt, Rinehart & Winston, Inc., 1964.

Lindquist, E. F. Design and Analysis of Experiments in Psychology and Education. Boston: Houghton Mifflin Co., 1953.

MacDonald, James B. Thoughts About Research in Schools. Educational Leadership. 23: 601-604; April, 1966.

McCall, W. A. How to Experiment in Education. New York: Macmillan Co., 1923.

Monroe, Walter S., and others. Ten Years of Educational Research, 1918-1927. Bulletin No. 42. Urbana: University of Illinois, 1928.

Mouly, George J. The Science of Educational Research. New York: American Book Co., 1963.

Nasatir, David. Global Rating Form. New York: Bureau of Applied Social Research, Columbia University, 1964. (multilith)

Perdew, Philip W. Criteria of Research in Educational History. Journal of Educational Research. 44: 217-233; November, 1950. (also Phi Delta Kappan. 32: 134-136; December, 1950).

Robinson, Helen M. Present and Future Research in Reading. In Education for Tomorrow -- Reading. The Pennsylvania State University, 1965. pp. 24-25.

Scates, Douglas E The Forming of Categories in Research. Journal of Educational Research. 29: 567-578; April, 1936.

Scates, Douglas E Assumptions Underlying Research Data. Journal of Educational Research. 34: 241-254; December, 1940.

Scates, Douglas E., and Hoban, Charles F., Jr. Critical Questions for the Evaluation of Research. Journal of Educational Research. 31: 241-254; December, 1937.

Shannon, J. R. Experiments in Education: A New Pattern and Frequency of Types. Journal of Educational Research. 48: 81-93; October, 1954.

Shaycroft, Marion F., and Altman, James W. A Procedure for Evaluating Graduate Research on the Basis of the Thesis. AIR Project NR 153-146. Pittsburgh: American Institutes for Research, 1955.

Stanley, Julian C. Quasi-Experimentation. School Review. 73: 197-205; Autumn, 1965.

Symonds, Percival M. A Research Checklist in Educational Psychology. Journal of Educational Psychology. 47: 100-109; February, 1956.

Tate, Merle W. Operationalism, Research, and a Science of Education. Harvard Educational Review. 20: 11-27; January, 1950.

Tatsuoka, Maurice M., and Tiedman, David V. Statistics as an Aspect of Scientific Method in Research on Teaching. In Handbook of Research on Teaching. (Edited by N. L. Gage.) Chicago: Rand McNally & Co., 1963. pp. 142-170.

Travers, Robert M. W. An Introduction to Educational Research. New York: Macmillan Co., 1964.

Tyler, Ralph W. What is Evaluation? Supplementary Educational Monographs, No. 88. Chicago: University of Chicago Press, 1958. pp. 4-9.

Van Dalen, D. B. A Research Check List in Education. Educational Administration and Supervision. 44: 174-181; May, 1958.

Wandt, Edwin. A Cross-Section of Educational Research. New York: David McKay Co., Inc., 1965.

Whitney, Frederick L. The Elements of Research. New York: Prentice-Hall, Inc., 1942.

Winer, B. J. Statistical Principles of Experimental Design. New York: McGraw Hill Book Co., 1962.

Wolfle, Daeì L., and others. Standards for Appraising Psychological Research. American Psychologist. 4: 320-328; August, 1949.

APPENDICES

APPENDIX A. Summary Tables on Related Literature

TABLE I

A LIST OF REVIEWS OF RESEARCH ON ELEMENTARY SCHOOL MATHEMATICS

Author	Year Published	Number of References Cited	Span of Years	Topic of Research
Smith	1918	---	-------	elementary school arithmetic
Woody	1931	68	1927-30	elementary school arithmetic
Brueckner	1934	33	1931-33	elementary school arithmetic
Woody	1935	96	1930-34	elementary school arithmetic
Brueckner et al.	1937	50	1935-37	elementary school arithmetic
Wilson	1941	115	1911-39	elementary school arithmetic
Brownell & Grossnickle	1942	142	1937-42	mathematics in grades 1-6
Brownell	1945	82	1942-45	mathematics in grades 1-6
Moser et al.	1948	17	1945-48	aims and purposes in teaching mathematics
Spitzer & Burch	1948	46	1945-48	methods and materials in teaching mathematics

TABLE I (continued)

Author	Year Published	Number of References Cited	Span of Years	Topic of Research
Wilson	1950	133	1904-47	elementary school arithmetic
Burch & Moser	1951	104	1948-51	mathematics in grades 1-8
Glennon & Hunnicutt	1952 1958	210	1891-1958	application to the classroom
Morton	1953	17	1925-53	application to the classroom
Gibb & Van Engen	1957	112	1951-56	elementary school mathematics
Buswell	1960	140	1902-56	elementary school mathematics
Spitzer & Burns	1961	89	1957-60	elementary school mathematics
Spitzer	1962	20	1918-62	application to the classroom
Weaver & Gibb	1964	63	1960-63	elementary school mathematics

TABLE II

A LIST OF TOPICAL SUMMARIES OF RESEARCH ON
ELEMENTARY SCHOOL MATHEMATICS

Author	Year Published	Number of References Cited	Span of Years	Topic of Research
Beatty	1950	67	1919-49	developing problem solving ability
Bernstein	1959	32	1929-56	remedial arithmetic
Brown, J. C.	1925	41	1902-24	significant conclusions
Brownell	1938	34	1912-37	readiness in primary grades
Brownell et al.	1941	60	1901-40	arithmetic in grades 1 and 2
Burns	1965	24	1919-59	influential research
Dawson & Ruddell	1955	6	1939-50	the case for meaning theory
Fehr	1950	25	1942-50	applications of specific research
Fisher	1912	38	1887-1911	reasoning
Gane	1962	14	1935-60	important studies
Gibb	1953	9	1940-50	methods of teaching arithmetic

TABLE II (continued)

Author	Year Published	Number of References Cited	Span of Years	Topic of Research
Greene	1930	29	1911-29	remedial and drill materials
Hartung	1957	16	1926-56	quotient estimation
Hightower	1954	17	1912-38	effect of instruction on achievement in fundamental operations
Johnson	1944	39	1922-42	problem solving
Knipp	1944	57	1911-40	comparisons of methods of teaching
Payne	1965	20	1960-64	modern programs
Pikal	1957	12	1941-55	arithmetic in upper elementary grades
Riess	1947	153	1836-1946	number readiness
Ruch & Mead	1930	9	1914-27	subtraction
Sherer	1953	43	1928-51	concept development
Stone	1932	21	1930	methods of teaching

TABLE II (continued)

Author	Year Published	Number of References Cited	Span of Years	Topic of Research
Van Engen	1948	12	1942-47	organization and learning of arithmetic
Weaver	1966	92	1953-66	differentiated instruction
Wheat	1941	6	1938-40	self-instruction
Willey	1942	19	1919-40	social values and uses
Wilson	1948	12	1924-47	social utility
Wrightstone	1952	27	1924-49	influence of research practice

TABLE III

A LIST OF BIBLIOGRAPHICAL LISTINGS OF RESEARCH ON ELEMENTARY SCHOOL MATHEMATICS

Author	Year Published	Number of References Cited	Span of Years	Topic of Research
Smith & Goldziher	1912	34	1900-12	titles from American and foreign periodicals
Buswell & Judd	1925	320	to 1925	summary by topics
Buswell	1926-1947	960	1925-47	annual summary of investigations and selected list of references
Buswell	1930	519	1893-1929	survey of the research: nature and findings
Monroe & Engelhart	1931	128	to 1930	critical summary related to teaching of arithmetic
Stretch	1941	100	1914-40	bibliography of selected references
Wheat	1945	43	1936-44	list of master's theses
Hartung	1948-1964	360	1948-64	selected list of references

TABLE III (continued)

Author	Year Published	Number of References Cited	Span of Years	Topic of Research
Schreiber	1948	7	1936-46	summary of studies
Van Engen	1950	34	1947-50	selected list of references
Brown, K.	1953-1955	---	1952-54	summary of investigations
Brown, K.	1953	57	1952	bibliography of doctoral dissertations
Brown, K.	1955	43	1953	summary of investigations
Gibb	1954	18	1944-54	selected bibliography
Weaver	1957	71	1951-56	six-year summary
Weaver	1958-1965	443	1957-65	annual summary
Hunnicutt & Iverson	1958	23	1927-49	selected studies
Brown, K.	1958	25	1955-56	biannual summary
Brown & Kinsella	1960	15	1957-58	biannual summary
Schaaf	1960	235	1941-58	annotated bibliography

TABLE III (continued)

Author	Year Published	Number of References Cited	Span of Years	Topic of Research
Weaver	1960	42	1941-60	bibliography of selected summaries and critical discussions
Summers	1961	118	1918-52	bibliography of doctoral dissertations
Summers & Stochl	1961	68	1950-60	bibliography of doctoral dissertations
Weaver	1961	--	1957-58	summary of studies
Brown & others	1963	26	1959-60	biannual summary
Summers	1963	15	1962	bibliography of doctoral dissertations
Brown & Abell	1965	--	----	biannual summary
Brown & Abell	1965	50	1965	summary of investigations
Summers & Hubrig	1965	15	1963	bibliography of doctoral dissertations
Burns & Dessart	1965	53	1964	summary of investigations

TABLE III (continued)

Author	Year Published	Number of References Cited	Span of Years	Topic of Research
Payne & Goodman	1965	over 2,000	1945-63	key-word bibliography
Weaver	1966	118	1965	annual summary
Burns & Dessart	1966	46	1965	summary of investigations

APPENDIX B. Tables of Data on Studies of Reliability of the Instrument for Evaluating Experimental Research Reports

TABLE IV

ANALYSIS OF VARIANCE: SUMMARY OF DATA FOR FIRST STUDY

Source	SS	df	MS
Between articles	1961.6	9	218.0
Masking (A)	128.2	1	128.2
Between articles within A (BwA)	1833.4	8	229.2
Within articles	390.3	20	19.5
Between judges (C)	40.5	2	20.3
A x C	65.2	2	32.6
C x BwA	284.6	16	17.8
Total	2351.9	29	

TABLE V

F RATIOS FOR THE FIRST STUDY

Source	Error term	F	p
A (masking: fixed)	BwA	.56	
BwA (articles: random)	pooled, C, AC, BwA	11.75	$p \leq .01$
C (judges: random)	BC	1.14	
A x C	BC	1.83	
C x BwA	(no error term)		

TABLE VI

ANALYSIS OF VARIANCE: SUMMARY OF DATA FOR SECOND STUDY

Source		SS	df	MS
Between articles		1581.98	9	175.78
Masking (A)		110.21	1	110.21
Between articles within A (BwA)		1471.77	8	183.97
Between judges		371.49	11	33.77
Experience (C)		69.01	1	69.01
Field (D)		130.21	1	130.21
Experience x Field (C x D)		1.00	1	1.00
Between judges (judges within CD) (EwCD)		171.27	8	21.41
Interaction: articles x judges		1040.12	99	10.51
A x C		7.00	1	7.00
A x D		18.40	1	18.40
A x C x D		14.03	1	14.03
C x BwA D x BwA CD x BwA A x EwCD EwCD x BwA	pooled residual	1000.69	96	10.42
Total		2993.59	119	

TABLE VII

F RATIOS FOR THE SECOND STUDY

Source	Error term	F	P
A (masking: fixed)	B	.60	
BwA (articles: random)	e*	17.66	$P \leq .01$
C (experience: fixed)	E	3.22	
D (field: fixed)	E	6.08	
EwCD (judges within experience by field: random)	e*	2.05	
C x D	E	.05	
A x C	e*	.67	
A x D	e*	1.77	
A x C x D	e*	1.35	

*pooled residual - 10.42

TABLE VIII

SUMMARY OF RELIABILITY COEFFICIENTS
FOR THE SECOND STUDY

Judges	N	Interrater agreement (AOV)	Individual reliability (Snedecor)
Elementary education faculty	3	.77	.53
Educational psychology faculty	3	.85	.65
Elementary education doctoral candidates	3	.92	.78
Educational psychology doctoral candidates	3	.78	.54
Total set of raters	12	.94	.57

APPENDIX C. Tables of Summaries Resulting from Categorization

TABLE IX

FREQUENCY OF REPORTS BY JOURNAL SOURCE

American Education	1
American Educational Research Journal	3
American Journal of Mental Deficiency	4
American Journal of Psychology	2
Arithmetic Teacher	158
Baltimore Bulletin of Education	1
California Journal of Educational Research	14
Catholic Education Review	1
Chicago Schools Journal	4
Child Development	23
Childhood Education	5
Education	8
Educational Administration and Supervision	9
Educational Method (Journal of Educational Method)	16
Educational Outlook	1
Educational Research Bulletin	27
Elementary English Review	2
Elementary School Journal (Elementary School Teacher)	132
Harvard Educational Review	2
High Points	1
Instructor	1
Journal of Applied Psychology	1
Journal of Education	2
Journal of Educational Psychology	57
Journal of Educational Research	138
Journal of Exceptional Children	2
Journal of Experimental Education	30
Journal of Experimental Psychology	2
Journal of Genetic Psychology (Pedagogical Seminary; Pedagogical Seminary and Journal of Genetic Psychology)	37
Journal of Psychology	1
Journal of Social Psychology	1
Mathematics Teacher	36

TABLE IX (continued)

Nation's Schools	1
National Educational Association Journal	2
National Elementary Principal	1
Ohio Schools	1
Peabody Journal of Education	11
Pittsburgh Schools	1
Pittsburgh University School of Education Journal	1
Reading Teacher	1
Scientific American	1
School Board Journal	1
School Executive	3
School Review	3
School Science and Mathematics	35
School and Society	5
Teachers College Record	4
Theory into Practice	1
Training School Bulletin	1
Wisconsin Journal of Education	4
	799

TABLE X

FREQUENCY OF REPORTS BY MATHEMATICAL TOPIC
AND TYPE OF STUDY

Type of Study	d	s	c	a	r	F	e	Total
Historical development and procedures (a-1)	2							2
Values of arithmetic (a-2)	2	8						10
Planning and organizing for teaching (a-3)	6	4		5		9	14	38
Attitude and climate (a-4)		13		1	5	2	3	24
Drill and practice (a-5a)	2	3	1	1			24	31
Problem solving (a-5b)	6	13		2	10	4	22	57
Estimation (a-5c)	1						1	2
Mental computation (a-5d)	3	1					5	9
Homework (a-5e)	1			1			4	6
Review (a-5f)							1	1
Checking (a-5g)							1	1
Writing and reading numerals (a-5h)		2					1	3
Foreign comparisons (a-6)	12	3			1	7		23
	35	47	1	10	16	22	76	207

TABLE X (continued)

Type of Study	d	s	c	a	r	F	e	Total
Pre-first-grade concepts (b-1)		18					1	19
Readiness (b-2)	1	5			2		2	10
Logical order (b-3)								0
Quantitative understanding (b-4)		5			2			7
Content to be included in grade (b-5)	7	9		2		3	3	24
Time allotment (b-6)		2				1		3
	8	39	0	2	4	4	6	63
Counting (c-1)		6	1					7
Number properties and relations (c-2)		3		1				4
Whole numbers (c-3)		2					2	4
Addition (c-3a)	4	4					8	16
Subtraction (c-3b)		6		1		3	6	16
Multiplication (c-3c)	1	4			1		3	9
Division (c-3d)	10	7		1		1	6	25
Fractions (c-4)		5		1			2	8
Addition (c-4a)						1		1

TABLE X (continued)

Type of Study	d	s	c	a	r	F	e	Total
Subtraction (c-4b)						1		1
Multiplication (c-4c)							2	2
Division (c-4d)		1				1	3	5
Decimals (c-5)		2			1		2	5
Percentage (c-6)	1	2		1			2	6
Ratio and performance (c-7)								0
Measurement (c-8)		13		1			9	23
Negative numbers (c-9)							1	1
Algebra (c-10)								0
Geometry (c-11)							3	3
Sets (c-12)		1		1				2
Logic (c-13)	1			3		1		5
Our numeration system (c-14)		2					1	3
Other numeration systems (c-15)				2			2	4
Probability and statistics (c-16)							4	4
	17	58	1	12	2	8	56	154

TABLE X (continued)

Type of Study	d	s	c	a	r	F	e	Total
Textbooks (d-1)	21	1				1	2	25
Workbooks (d-2)	1						1	2
Manipulative devices (d-3)				2		1	9	12
Audio-visual devices (d-4)	2	1		1			4	8
Programmed instruction (d-5)		1		1		1	8	11
Readability and vocabulary (d-6)	6	5				1	3	15
Quantitative concepts in other subject areas (d-7)	2	2					1	5
	32	10	0	4	0	4	28	78
Diagnosis (e-1)	2	19	2	3		1	4	31
Remediation (e-2)	2	1	12	9	2	1	8	35
Enrichment (e-3)	1	2	1			3	8	15
Grouping procedures (e-4)	2			12	1	3	10	28
Physical, psychological, and/or social characteristics (e-5)		7	1		1	1		10
Sex differences (e-6)					1	7		8
Socio-economic differences (e-7)		3					1	4
	7	32	16	24	25	16	31	131

TABLE X. (continued)

Type of Study	d	s	c	a	r	F	e	Total
Testing (f-1)	3	7		4	4		5	23
Achievement evaluation (f-2)		17		1	3	12	1	34
Relation to achievement (f-3)		2			3	2		7
Age (f-3a)					3	3		6
Intelligence (f-3b)	1	1			15	3		20
Effect of parental knowledge (f-4)		1					2	3
Effect of teacher background (f-5)		1			1		4	6
	4	29	0	5	29	20	12	99
Transfer (g-1)		1					12	13
Retention (g-2)		3		2			7	12
Generalization (g-3)		1				1		2
Organization (g-4)	2	1					2	5
Motivation (g-5)				3		1	4	8
Piagetian concepts (g-6)	2	9		1		3	9	24
Reinforcement (g-7)							3	3
	4	15	0	6	0	5	37	67

TABLE XI

FREQUENCY OF QUALITATIVE VALUE BY YEARS

	1900-29	1930-50	1951-65		1900-29	1930-50	1951-65
13			2	29		5	3
14			2	30		5	3
15			6	31	1	3	6
16			2	32	3	4	6
17		1	4	33	2	5	3
18			5	34	1	5	4
19			6	35		6	2
20			6	36	4	1	1
21		2	8	37	5	4	2
22		1	7	38	6	3	1
23			12	39	4	4	
24	1	2	7	40	4	3	1
25		1	3	41	7	3	1
26		1	7	42	6	2	
27		2	6	43	2	3	
28	1	1	15	44	1		

APPENDIX D. Categories and Coding for Mathematical Topic

a. Educational objectives and instructional procedures
   1) Historical development and procedures
   2) Values of arithmetic
   3) Planning and organizing for teaching (meaning approach; multi-graded; departmentalized, self-contained, non-graded; team teaching; modern, traditional; exposition, discovery; incidental, systematic; activity program; teaching practices)
   4) Attitude and climate
   5) Specific procedures
      a) Drill and practice
      b) Problem solving
      c) Estimation
      d) Mental computation
      e) Homework
      f) Review
      g) Checking
      h) Writing and reading numerals
   6) Foreign comparisons

b. Topical placement
   1) Pre-first-grade concepts
   2) Readiness
   3) Logical order
   4) Quantitative understanding
   5) Content to be included in grade
   6) Time allotment

c. Basic concepts (and methods of teaching them)
   1) Counting
   2) Number properties and relations
   3) Whole numbers
      a) Addition
      b) Subtraction
      c) Multiplication
      d) Division
   4) Fractions
      a) Addition
      b) Subtraction
      c) Multiplication
      d) Division
   5) Decimals
   6) Percentage
   7) Ratio and proportion
   8) Measurement (time, denominate numbers)
   9) Negative numbers (integers)
   10) Algebra
   11) Geometry

12) Sets
13) Logic
14) Our numeration system
15) Other numeration systems
16) Probability and statistics (graphing)

d. Materials
1) Textbooks
2) Workbooks
3) Manipulative devices
4) Audio-visual devices
5) Programmed instruction
6) Readability and vocabulary
7) Quantitative concepts in other subject areas

e. Individual differences
1) Diagnosis (errors)
2) Remediation (slow learner, underachiever)
3) Enrichment (acceleration)
4) Grouping procedures (ability, homogeneous, individualized, flexible)
5) Physical, psychological, and/or social characteristics
6) Sex differences
7) Socio-economic differences

f. Evaluating progress
1) Testing
2) Achievement evaluation
3) Relation to achievement
a) Age
b) Intelligence
4) Effect of parental knowledge
5) Effect of teacher background

g. Studies related to learning theory
1) Transfer
2) Retention (retroactive inhibition)
3) Generalization
4) Organization (process, reasoning)
5) Motivation
6) Piagetian concepts
7) Reinforcement (knowledge of results)

APPENDIX E. Categories and Coding for Type of Study

d	Descriptive:	research in which the researcher reports on records which may have been kept by someone else; includes reviews, historical studies, and textbook analyses or comparisons
s	Survey:	research which attempts to find characteristics of a population by asking a sample through the use of a questionnaire or interview; includes also the status study, in which a group is investigated as it is to ascertain pertinent characteristics (measures assigned variable only)
c	Case study:	research in which the researcher describes in depth what is happening to one designated unit, usually one child
a	Action research:	research which uses nominal controls; generally teacher or school originated; procedures of actual practice may be described
r	Correlational:	research which studies relationships between or among two or more variables; uses correlational statistic primarily
F	Ex post facto:	research in which the independent variable or variables were manipulated in the past; the researcher starts with the observation of a dependent variable or variables. He then studies the independent variables in retrospect for their possible effects on the dependent variables. (He may examine interrelationships of two or more assigned variables or two or more levels of one assigned variable)
e	Experimental:	research in which the independent variable or variables are manipulated by the researcher to quantitatively measure their effect on some dependent variable or variables, to test a logically derived hypothesis

APPENDIX F. Instrument for Evaluating Experimental Research Reports

Directions:

Evaluate with the nine underlined questions which follow. The quality of the research report in terms of each question should be rated on a five-point scale. The specifications for these five points are:

1) Excellent: all requirements for the question are met; nothing essential could be added

2) Very good: most requirements are met

3) Good: some requirements are met

4) Fair: a few requirements are met

5) Poor: none or too few of the requirements are met

Certain "key points" should be considered in ascertaining a rating for each question. These are listed below the question, followed by adjectives which indicate the continuum on which the "key point" should be assessed. Do NOT make a response to these "key points." They are intended to focus the attention of all raters on the same pertinent aspects of each question.

Please make only nine responses for each article, one for each question.

Instrument for Evaluating Experimental Research Reports

Marilyn N. Suydam
The Pennsylvania State University

1. How practically or theoretically significant is the problem? (1-2-3-4-5)

a. Purpose	(important---non-important)
b. Problem origin	
1) Rationale	(logical---illogical)
2) Previous research	(appropriate---inappropriate)

2. How clearly defined is the problem? (1-2-3-4-5)

a. Question	(operational---vague)
b. Hypothesis(es)	(relevant---irrelevant)
	(logical---illogical)
c. Independent variable(s)	(relevant---irrelevant)
d. Dependent variable(s)	(operational---vague)
	(relevant---irrelevant)

3. How well does the design answer the research question? (1-2-3-4-5)

a. Paradigm	(appropriate---inappropriate)
b. Hypothesis(es)	(testable---untestable)
c. Procedures	(clear---unclear)
d. Treatments	(replicable---unreplicable)
	(appropriate---inappropriate)
e. Duration	(appropriate---inappropriate)

4. How adequately does the design control variables? (1-2-3-4-5)

a. Independent variable(s)	(uncontaminated---contaminated)
b. Administration of treatment	(rigorous---unrigorous)
c. Teacher or group factors	(controlled---uncontrolled)
d. Subject or experimenter bias	(controlled---uncontrolled)
e. Halo effect	(controlled---uncontrolled)
f. Extraneous factors	(controlled---uncontrolled)
g. Individual factors	(controlled---uncontrolled)

5. How properly is the sample selected for the design and purpose of the research? (1-2-3-4-5)

a. Population	(appropriate---inappropriate)
b. Drawing of sample	(random---unspecified)
c. Assignment of treatment	(random---unspecified)

d. Size (appropriate---inappropriate)
e. Characteristics (appropriate---inappropriate)

6. How valid and reliable are the measuring instruments or observational techniques? (1-2-3-4-5)

a. Instrument or technique
   1) Description (excellent---poor)
   2) Validity (appropriate---inappropriate)
   3) Reliability for population (excellent---poor)
b. Procedure of data collection (careful---careless)

7. How valid are the techniques of analysis of data? (1-2-3-4-5)

a. Statistical tests
   1) Basic assumptions (satisfied---unclear)
   2) Relation to design (appropriate---inappropriate)
b. Data
   1) Treatment (appropriate---inappropriate)
   2) Presentation (clear---unclear)
   3) Level of significance (appropriate---inappropriate)
   (specified---unspecified)
   4) Discussion (accurate---inaccurate)

8. How appropriate are the interpretations and generalizations from the data? (1-2-3-4-5)

a. Consistency with results (excellent---poor)
b. Generalizations (reasonable---exaggerated)
c. Implications (reasonable---exaggerated)
d. Limitations (noted---not noted)

9. How adequately is the research reported? (1-2-3-4-5)

a. Organization (excellent---poor)
b. Style (clear---vague)
c. Grammar (good---poor)
d. Completeness (excellent---poor)
(replicable---unreplicable)